

## Think of a number and double it

**Even with the US federal budget in surplus, the scientific community needs to present a more sophisticated argument than merely demanding an across-the-board doubling of research funding for government agencies.**

Scientific societies in the United States are working to muster support in the Senate for a bill — the National Research Investment Act, or S1305 — that would double civilian research and development spending, from \$34 billion to \$68 billion a year, over the next ten years. The societies view the bill not merely as a desirable goal, but as a way of drawing together science's fragmented base of political support. But the measure is unlikely to pass into law, having attracted only 15 supporters since it was put forward last October by two Republican and two Democrat senators (see page 4).

One of S1305's greatest strengths is its simplicity. By increasing the level of support at all science agencies at once, it avoids painful and unseemly arguments about which deserve most support. This has enabled more than a hundred groups representing the scientific and engineering communities to unite behind the measure to an unprecedented degree, and to deliver to the senators a straightforward message in the name of some three million scientists and engineers.

But this lack of sophistication may also be the measure's undoing. The arguments for doubling resources for each of the US federal government's research and development activities are, to put it politely, undeveloped. Indeed there are some areas of government research activity that would not deserve more money even if the citizenry was queueing up at the laboratory gates to deposit more tax dollars directly.

Furthermore, if the National Research Investment Act were to gather sufficient support to reach the Senate floor, the sheer amount of money involved — \$170 billion in extra spending over ten years — would, as Senator Jay Rockefeller (Democrat, West Virginia) has observed, smoke out enemies that science never even knew it had.

At heart, therefore, S1305 may be well-intentioned, but it is also flawed. It attempts to deny a rational science policy by implying that both the research enterprise and society stand to benefit from a large and indiscriminate increase in current civilian research spending. But it ill becomes the community to argue against an intelligent and highly

selective science policy. As often stated during the recent years of budgetary restraint, US science needs more honest and more vigorous identification of priorities. It would be the height of folly for the community to turn its back on this requirement the moment a balanced federal budget is in sight.

The message of S1305 is simple and easy to explain: "spend". This is already a distressingly familiar refrain in Washington, where major special interest groups, notably the military and the roads lobby, are preparing for a full-frontal assault on the spending limits agreed in last year's balanced budget between President Bill Clinton and the Congress. These interest groups carry heavier artillery than the science community can ever hope to muster.

Some argue that authorization bills — which set theoretical limits on future government spending — such as S1305 are unimportant, as real spending decisions are taken by appropriations committees on an annual basis. But if the Senate and the House of Representatives were to confront some jurisdictional issues and actually pass authorization bills for science agencies, these would indeed carry influence, as they already do in other parts of the government. It is therefore to be hoped that senators such as Bill Frist (Republican, Tennessee), who have expressed reservations about S1305, will work with its sponsors to develop a more realistic alternative, and then work with colleagues in the House of Representatives to pass it.

The community's top priority this year is to extend the aura that surrounds the National Institutes of Health. Considerable progress has already been made in educating Congress about the importance of the National Science Foundation in supporting health and other government missions. Other such connections, including the role of Department of Energy facilities in supporting the life sciences, remain to be better understood and articulated (see page 3). Science will ultimately be sold to the Congress only on the intelligent explanation of each part of the research enterprise. □

## Challenges ahead

**The success of Britain's Technology Foresight exercise is no grounds for complacency about its successor.**

Does good science inexorably lead to new products and economic growth (the 'science push' model)? Or conversely, does a dynamic industrial base act as the spur for novel scientific ideas ('technology pull')? For the past five years, Britain's Technology Foresight exercise has attempted to finesse this conundrum by suggesting a third approach, namely that innovation and growth emerge from a complex interaction between science and the market place, involving close communication and substantial feedback between actors in the two separate fields.

Certainly the apparent success of the exercise, indicated by the British government's recent announcement that it is launching a second programme, called merely Foresight (see page 8), has confounded sceptics who initially complained that it would inevitably place a strait-jacket on British science, and degenerate into a doomed attempt to pick technology 'winners'. Success is based on the way Technology Foresight has become a device for injecting a user dimension

into debates on scientific priorities that has proved acceptable to political parties of both left and right. At its best, Technology Foresight can represent the 'bottom-up' approach to setting research priorities that fits well with the flexible, competitive economies of the late 1990s. It is perhaps not surprising that similar techniques are now being explored in countries as diverse as India and Germany (see page 7).

But Britain must not rest on its laurels. There are still plenty of problems with the implementation of both Technology Foresight and its successor, such as a failure to penetrate sufficiently the small and medium-sized businesses that represent the most innovative sector of the economy. The effectiveness of targeting Foresight activities around broad social goals, as proposed for the second exercise, has yet to be proved. Finally, careful prioritization is no alternative to the adequate funding of the science base. The cultural change in British science over the past five years means that additional funding is now more important than ever — not that it is no longer required. □



# US seeks fair deal for biologists on synchrotron source access

[WASHINGTON] US government agencies are considering how the management and financing of synchrotron radiation sources can best be adapted to cope with the growing number of biologists who want to use them.

A working group of senior officials has been set up by Arthur Bienenstock, associate director of basic research at the White House Office of Science and Technology Policy and former director of the Stanford Synchrotron Radiation Light Source in California. The science office expects the informal group to come up with solutions, which it promises to implement.

The move comes partly in response to complaints from structural biologists that access to the synchrotron sources can take six months or more to arrange. According to officials, the working group will assess likely future demands on the facilities from life scientists, and recommend how to meet needs for instruments, staff and support infrastructure.

Synchrotrons are huge, circular electron beams emitting intense X-rays which researchers use to study crystal structures. Four out of seven of the US facilities are operated by the Department of Energy (DOE) at an annual cost of almost \$180 million. But structural biologists — most supported by the National Institutes of Health (NIH) — now account for one-third of the users, and the proportion is growing sharply.

Scientists hope that better coordination between the agencies will help to persuade the Congress to relieve pressure on the energy department's budget for operating the synchrotrons by emphasizing their importance to health research. It could also persuade the NIH, whose financial outlook is far healthier than DOE's, to do more to support its researchers at DOE facilities.

The group, led by Marvin Cassman, director of the National Institute of General Medical Sciences at NIH, met for the first



Source of dissatisfaction: Brookhaven's synchrotron source is, like others, in heavy demand.

time three weeks ago. Its other members are Patricia Dehmer, head of basic energy sciences at DOE; Ari Patrinos, head of biological and environmental research at DOE; Judy Vaitukaitis, director of the NIH's National Center for Research Resources; Mary Clutter, associate director for life sciences at the National Science Foundation (NSF); and Jack Rush of the National Institute of Standards and Technology (NIST).

Last week, Cassman asked the DOE's Biological and Environmental Research Advisory Council (BERAC) to summarize available data on the outlook for structural biologists who use DOE synchrotrons. He told the panel: "The people with the money lock up as much beam time as they can, and it gets harder and harder for general users to get on."

The largest DOE synchrotron facilities involved are at Stanford, the Advanced Photon Source at the Argonne National Laboratory in Illinois, and the National Synchrotron Light Source at the Brookhaven National Laboratory, New York. NIST and NSF run smaller facilities.

One issue the working group is expected to tackle is recurrent complaints about the alleged exclusion of individual investigators from beamlines — experimental stations on

the synchrotrons — paid for by wealthy consortia of researchers and corporations.

Each DOE facility was built by the department, and costs it tens of millions of dollars to operate a year. Additional beamlines, which tap X-rays from the synchrotron, are often equipped by private consortia of universities and corporations, who agree to allocate 25 or 30 per cent of the available time on their beamline to other government-funded scientists.

But in practice this allocation is not always available. According to Vaitukaitis, some consortia do not provide the support staff that occasional users need. The question of who pays for these staff, for computers and other infrastructure, and even for buildings to house them, is often unresolved.

Such resource questions are compounded by the way the biologists operate, which contrasts with that of the synchrotrons' traditional user base — physicists and materials scientists. They did experiments that ran for weeks, and took a strong interest in designing, building and supporting the facilities.

The new users have a different approach, says Denis McWhan, associate director of the Brookhaven laboratory. "They don't know how a beamline works, and they want to come in with their crystal and walk out with their data the same day," he says. "We don't have the people to provide the support they need."

Government agencies have started to provide such support: last year, DOE spent \$8 million on it and the NIH \$6 million. But an extra \$5.5 million for support scientists and technicians, and a capital investment of \$12 million, could double the beam time available for structural biologists, according to an unpublished report by Keith Hodgson, professor of chemistry at Stanford University and chairman of BERAC, and Ed Lattman, chairman of BERAC, at Johns Hopkins University in Baltimore, Maryland. **Colin Macilwain**

## Russian reshuffle places Bulgak at helm

[MOSCOW] Vladimir Bulgak, formerly a deputy prime minister whose responsibilities included science and technology, was made minister of science and technology in last week's cabinet reshuffle by Russian President Boris Yeltsin.

Unlike Vladimir Fortov, who lost his position in the cabinet as minister of science and technology policy, Bulgak is not a member of the Russian Academy of Sciences. That should reduce the possibility of a

conflict of interests in dividing government funding between different scientific organizations.

Bulgak has been an active supporter of reforms to the academy, and has rejected demands from scientific labour unions that such reforms should be postponed (see *Nature* 388, 315 & 390, 328; 1997).

He had been considered as a possible candidate to succeed former prime minister Victor Chernomyrdin. **Carl Levitin**

# Animal deaths turn shuttle into 'necrolab'

[WASHINGTON] An unexpectedly high mortality rate among research animals in orbit has complicated a decision by the US space agency on whether to re-fly the Neurolab space shuttle mission this autumn.

Managers at the National Aeronautics and Space Administration (NASA) expect to decide this week if the Neurolab experiments will be repeated to give researchers more data on how the brain and nervous system adapt to spaceflight (see *Nature* 392, 851; 1998). Before that, however, they would like to understand what caused 57 of 96 newborn rats to die during the two-week flight.

Six of Neurolab's 26 experiments used eight-day-old rats as test subjects to enable researchers to study the effects of weightlessness during the earliest stages of development. But most of the newborn rodents were unable to feed properly from their mothers, became malnourished and died, despite astronauts' attempts to keep them alive.

The high mortality rate disappointed and puzzled mission planners, who thought they had solved a similar problem that occurred on a previous shuttle mission. On that flight, some 60 per cent of five-day-old rats died, but only five per cent of the eight-day-olds were lost.

A NASA Animal Care and Use Committee ruled that only eight-day-olds would be allowed on Neurolab, but now they, too, have suffered high casualties. NASA's chief veterinarian, Joseph Bielitzki, points out that 14-day-old and adult rats experienced no such problems on the recent flight. "I wish I could



Lost in space: of the snails' eggs on Neurolab, only half the number expected hatched in orbit.

tell you what the big difference was," he says.

Bielitzki says the youngest rats "weren't being confined to the nest adequately" in weightlessness, and had trouble floating back to their mothers inside their Research Animal Holding Facility. A different holding unit, the Animal Enclosure Module, housed eight-day-olds on their successful shuttle flight.

Both enclosures are built by NASA's Ames Research Center, and both have about the same volume, but their configurations are different. The enclosure design is one of the areas NASA needs to take a look at, says Bielitzki.

Rats were not the only animals to die unexpectedly on Neurolab. All but 25 of 225 swordtail fish onboard perished when temperatures inside their specialized aquarium become too high. An air-intake line seems to have become clogged, perhaps because the busy astronauts were unable to clean it as often as required.

Less than half the expected number of snail eggs hatched in orbit, and two of four adult oyster toadfish were dead on arrival back on Earth.

Neurolab's research teams were scrambling early this week to make the best of the situation, and most of the scientists say they will be able to meet many of their research goals. But the shortage of test animals has forced teams to share body parts and tissue samples.

In the case of the swordtails, for example, the principal investigator will have just enough samples to meet his own needs. But co-investigators will go wanting.

The experience shows that keeping research animals healthy in space remains an uncertain business. A panel is being formed by NASA and its main partner in the Neurolab project, the National Institutes of Health, to look into why the animals died and "what we need to do to get on with developmental biology in space", says Bielitzki.

Meanwhile, NASA must decide whether to re-fly Neurolab, with or without the eight-day-old rats. The verdict depends in part on US and Russian plans for building the space station, which was to have begun this summer but will now be delayed. **Tony Reichhardt**

## US plan to double spend comes under fire as 'unrealistic exercise'

[WASHINGTON] An important subcommittee of the US Senate has expressed strong reservations about a plan to double civilian research and development spending over the next ten years, casting yet further doubt on the chances that the ambitious proposal will pass into law.

Senator Bill Frist (Republican, Tennessee), chair of the science, technology and space subcommittee of the Senate Commerce Committee, said after a hearing on the proposed law last week that it was "likely" that his committee would draft an alternative proposal.

The proposed National Research Investment Act, or S1305, is backed by 15 senators, and would set spending limits for all major science agencies that would allow their funding to double over the next ten years. These limits would be reached only if appropriations committees, which exercise the real budgetary power in the Congress, chose to reach them.

Frist says he would prefer a bill that

could gather the signatures of 30 or 40 senators and could actually be considered on the Senate floor. "We want to do something that will not be a useless exercise, something that is more than symbolic," he says. Frist adds that the numbers in his bill may be "tied to indices", such as the rate of inflation or growth in national product.

Senator Jay Rockefeller, the senior Democrat on the subcommittee, shares Frist's feelings. "We'll draw up a bill that has a greater chance of passing," he says. However, observers think there is little chance of either S1305 or an alternative measure passing the Senate this year.

Senators Phil Gramm (Republican, Texas), Jeff Bingaman (Democrat, New Mexico), Joseph Lieberman (Democrat, Connecticut) and Arlen Specter (Republican, Pennsylvania) all testified before the subcommittee in favour of S1305. Pete Domenici (Republican, New Mexico), chairman of the powerful Budget Committee, who also supports the measure, disappointed

his allies by not showing up to testify.

The main point being pushed by the bill's advocates is that it deserves support because its "doubling over ten years" message is, in Lieberman's words, "simple, easily understandable and easy to communicate".

Science lobbyists still hope that Frist and Rockefeller will talk to S1305's supporters and eventually come round to backing a similar measure. In the meantime, they see the proposal as a useful rallying cry for science funding.

But George Brown (Democrat, California), senior Democrat on the House of Representatives Science Committee, made his fiercest attack yet on the measure in addressing a meeting last week of the American Association for the Advancement of Science. "It's a basic scam by a number of senators," he said. "It ain't going to happen, it's impossible; [its supporters] aim to get a lot of good press releases out of it and then forget it after the elections." **Colin Macilwain**



# Merger plans worry Japan's industrialists

[TOKYO] Japanese industrialists are voicing scepticism about the country's plans to restructure its science-related ministries, which otherwise have faced little public or political opposition. Many doubt whether the restructuring will increase efficiency as promised and thus genuinely improve the country's research base.

The government announced last November that it plans to merge its Science and Technology Agency (STA) with the education ministry by 2001 (see *Nature* 390, 327; 1997), as part of a series of reforms designed to improve the country's administration.

The hope is that the merger, which follows the 1996 Basic Science and Technology Law committing Japan in principle to increasing spending on science and technology by 50 per cent by the year 2001, will help to invigorate Japan's scientific research base.

But senior research managers at some of Japan's leading research-orientated companies are concerned that a culture clash between the traditional Ministry of Education, Science, Sports and Culture (Monbusho) and STA may delay reforms so much that the government is unable to deliver its promised efficiency gains.

Michiyuki Uenohara, for example, chairman of the NEC Research Institute, points out that Monbusho and STA have significantly different cultures. Monbusho has a reputation for being one of Japan's most conservative ministries, handing out research funds widely and evenly.

In contrast, STA prefers a more centralized, selective approach, and has been trying to introduce research assessment and other modern research management techniques to Japan. "There will be constant fighting," predicts Uenohara.

Susumu Nishimura, senior executive director with Banyu Tsukuba Research Institute, a leading Japanese pharmaceutical company, worries that a merger will be insufficient to produce the desired improvements unless the attitude of officials changes.

Such industrialists argue that the real need is to improve links between universities and industry, and to have more flexible research budgets, greater government transparency and less waste. They argue that the merger will not actually reduce waste, and may inadvertently increase bureaucracy.

For example, the ministry of international trade and industry and the ministry of agriculture, as well as Monbusho and STA, all support various kinds of genome-related research. "The overlap is often redundant," says Nishimura. But he points out that much overlap will remain even after the merger.

The additional funding brought about by the Basic Science and Technology Law, as well as various supplementary budgets, may

be stimulating the economy and research in general terms, he says, but there is still too much waste.

Despite the new law, change in Japanese research has been slow, and collaboration between industry and universities is still much less common than in the United States, says Uenohara. His company, a world leader in research into carbon nanotubes and fullerenes, is unlikely in the near future to reverse its current policy of conducting most of its fundamental research in Princeton, New Jersey.

Other critics of the merger are concerned that the education ministry's resources are mainly dedicated to primary and secondary education, and that its higher education bureau is "very poorly managed". It also has a poor record in fostering links between universities and industry.

Supporters of the merger argue that it will strengthen science in higher education, and encourage collaboration between Mon-

busho- and STA-run agencies — for example between Monbusho's Institute of Space and Astronautical Science and STA's National Space Development Agency.

The merger will, in principle, remove bureaucratic boundaries and make it easier for administrators to work more closely together, as they will all work for the same ministry, says one Monbusho official.

But many industrialists argue that integrating STA into Monbusho could lead to a situation in which inefficiency remains while research becomes a lower government priority, as there will no longer be an agency entirely dedicated to science and technology.

Senior Japanese academics have also expressed concern. They worry that the merger may lead to a centralized science administration that will damage the culture of basic research in Japan if it fails to encourage a more 'bottom-up' approach (see *Nature* 391, 431; 1998). The plan has still to be approved by Japan's parliament.

Richard Nathan

## Economy lifeline viewed as mixed blessing

[TOKYO] Japan's largest-ever economic stimulus package, unveiled two weeks ago and intended to halt the deterioration in the country's economy, has met with mixed reactions from the scientific community.

Part of the extra money to be spent on public works — which includes projects linked to environmental protection, energy, information and telecommunications — will go on improving research facilities at nationally run universities and institutes.

One beneficiary will be the University of Tokyo, which ran into difficulties last year when it failed to obtain government funding to acquire the land for its new campus in Kashiwa, Chiba prefecture (see *Nature* 392, 429; 1998). The stimulus package will now allow the university to transfer its Institute of Solid State Physics to the new campus.

Several projects will receive additional funding from the stimulus package as part of its commitment to promoting 'commercially applicable research'. For

example, the Biomolecular Frontier Programme and the Brain Research Institute will receive a further ¥3.4 billion (US\$26 million) and ¥5.3 billion respectively towards the completion of new research buildings, and ¥18.7 billion will go to work on the Japanese module of the International Space Station.

But many scientists remain worried that the hastily compiled package risks wasting money by investing heavily in expensive equipment and buildings without ensuring their efficient use.

A director of one research institute run by the Ministry of Education, Science, Sports and Culture says: "It seems that the construction, transportation and telecommunications ministries, as well as related industries, managed to win the money-grabbing game. There has been very little planning behind this."

His major concern is that, if the new stimulus package were to be followed by a sharp recession in a couple of years' time, universities and research institutes could

be left with pristine facilities and state-of-the-art equipment, but no money to run them.

Almost half of the ¥16.65 trillion package will be spent on public works projects proposed by national and local governments. Most of the public works funds will be spent on traditional projects such as reclamation and building highways, roads and bridges, but ¥1.5 trillion will be invested in 'new types' of projects such as developing a fibre-optic communications network.

The heavy emphasis in the stimulus package on traditional public works is intended, it seems, to allow these areas to acquire what they failed to gain in the main 1998 budget that came into effect on 1 April.

This budget had originally been drawn up under new fiscal legislation, introduced last November, to cap major spending and reduce the national deficit. But the government has since amended the law to allow it to launch the economy-boosting package.

Asako Saegusa



**Sun spotter:** heavy-water detector could provide clues to the mass density of the Universe.

## Scientists go down a mine to observe the Sun's neutrinos

[MONTREAL] Scientists and government representatives from many countries gathered last week near the northern Ontario mining town of Sudbury for the opening of the Sudbury Neutrino Observatory.

The observatory, in a nickel mine 2 km deep, consists of 9,500 light sensors surrounding a 12-metre diameter acrylic flask that will hold 1,000 tonnes of heavy water. The heavy water comes from Atomic Energy of Canada, which had a surplus of the expensive material as a result of a decline in sales of the CANDU reactor in the 1980s.

Interactions between neutrinos and the deuterium in the heavy water will produce light flashes. The observatory's detector will make an independent measurement of all three types of neutrinos (electron, tau and muon), so should be able to determine whether neutrinos produced inside the Sun change type as they travel outwards. If they do, they must have mass, and may make a substantial contribution to the Universe's total mass density.

**David Spurgeon**

## Bid to give legal protection to laboratory mice in US

[WASHINGTON] The US government is facing a renewed call to extend legal protection to some 20 million laboratory mice and rats not covered by current law.

The Animal Welfare Act, which is administered by the US Department of Agriculture (USDA), covers animals such as dogs, cats, rabbits, guinea pigs and non-human primates but, under an exemption approved in 1976, does not apply to mice, rats and birds.

Mice and rats are estimated to make up at least 85 per cent of US research animals. "These animals deserve the same protection as others used in experiments," said Tina Nelson, the executive director of the American Anti-Vivisection Society (AAVS). The society, a sharp critic of the treatment of laboratory animals, announced last week that it was petitioning the USDA. The move follows a previous unsuccessful attempt by other groups in 1989. "These animals are not afforded even minimal coverage by the Act," said Nelson.

Nelson argues that this has removed incentives for scientists to seek alternatives. She points, for example, to the "ascites method" used to produce monoclonal antibodies, in which tumours are grown in the abdomens of mice and the antibodies generated are tapped by syringe.

The act allows for unannounced laboratory inspections by USDA, and requires researchers to report annually on the numbers of laboratory animals in their care. Also, institutional committees that review experimental protocols must require investigators to consider non-animal alternatives.

The USDA has 180 days to respond to the petition. If it denies it, Nelson says, the AAVS will sue the government. The 1989 petition, which was similar to the current one, was denied by USDA. The petitioners sued, but a court found that they had not demonstrated

sufficient legal standing or injury to the rodents to justify the suit. The AAVS says its petition addresses those weaknesses.

Scientific groups, although not opposing the extension of the law in principle, say that the success of the petition could cripple the resource-starved USDA inspectorate, particularly as it would have to inspect small institutions such as community colleges and even high schools that use only rodents in their laboratories. "It would destroy USDA's already limited ability to meet its inspection mandates," says Tony Mazzaschi, a research policy analyst at the Association of American Medical Colleges.

Ralph Dell, director of the Institute for Laboratory Animal Research — a division of the National Research Council — estimates that the increased workload would require a doubling of USDA's \$9.2 million budget for enforcing the act.

Others argue that rodents are already cared for scrupulously. "The personnel at research facilities don't need a federal law in order to be motivated to care for their animals well. Good science demands it," says Barbara Rich, executive vice-president of the National Association for Biomedical Research.

The scientists also claim that extension of the law would make little difference to the vast majority of rodent-using institutions, which are already governed by other federal policies that cover rodent use.

For instance, scientists funded by the Public Health Service — including all those on National Institutes of Health grants — are covered by a policy on the humane use of laboratory animals that applies to all vertebrates. This policy does not, however, have the force of law, and it requires neither reporting of numbers nor surprise laboratory inspections.

**Meredith Wadman**

## Space researchers were not spying, says India's Supreme Court

[NEW DELHI] India's Supreme Court exonerated two leading space scientists and four others last week in the culmination of a spy case that rocked the country four years ago (see *Nature* 372, 491; 1994). The two scientists worked for the Indian Space Research Organization (ISRO).

The court also criticized the government of the state of Kerala for trying to reopen the investigation after it had been closed by a central government agency. The court ordered the state to pay \$2,500 in compensation to each of the six accused.

Nambi Narayanan and S. Chandrasekar

were scientists at ISRO's liquid propulsion centre in Thiruvandapuram when they were arrested by Kerala police and charged with passing information about the development of India's cryogenic rocket engine.

Also arrested were two women from the Maldives, and two Indian businessmen who had links with a Russian aerospace company. All were accused of being part of a spy ring financed by a third country.

The charges were based on the fact that addresses and contact numbers of ISRO scientists and of the businessmen were found in the hands of the Maldivian women,

one of whom worked for the islands' intelligence agency.

India's Central Bureau of Investigations, which took over the case, found no evidence of espionage and declared the case closed in April 1996. But the Kerala government disagreed and sought a new investigation by the state police.

Although it is not officially admitted, it is believed that low morale at ISRO caused by the investigation into the spying allegations has helped push India's geostationary launch-vehicle programme back by two years.

**K. S. Jayaraman**

# Foresight study blazes trail in Germany

[MUNICH] Materials and life scientists in Germany should work more closely together, according to the country's first pilot exercise in technology foresight. It also concludes that there should be more interdisciplinary projects with a wide range of potential applications.

The report of the exercise, published last week, identifies the most promising areas of research with potential applications as molecular architecture, molecular- and bioelectronics, and "materials whose properties are determined by interfaces".

The pilot study was carried out by a group headed by Dagmar Schipanski, an electronics engineer and former president of the Wissenschaftsrat, Germany's science council. The study involved 12 scientists selected by Germany's six major research organizations, including the Deutsche Forschungsgemeinschaft (DFG), the Fraunhofer Society and the Max Planck Society (MPS).

These organizations will now discuss the study's methods and its results. A decision about whether its approach to technology foresight should be expanded to other areas of research — possibly the life sciences — is expected in the next few months.

The study panel, working with German and international scientists including Nobel prizewinners Klaus von Klitzing of the Max Planck Institute for Solid State Physics in Stuttgart and Jean-Marie Lehn of the Louis

Pasteur University in Strasbourg, France, tried to predict areas of materials sciences likely to become particularly significant.

Although no direct funding recommendations are made, the authors specify research topics they consider as medium-term priorities. These include the synthesis and structural analysis of functional molecules with specified electrical, optical and ionic qualities, and investigations of the organization of very large self-organizing systems, and of protein folding.

The pilot study was set up by the Wissenschaftsrat which, after lengthy discussions, recommended four years ago that independent technology foresight studies of scientific research should be carried out in Germany. Materials science was chosen as the subject because the Wissenschaftsrat had already carried out a preliminary analysis of non-university materials science in Germany in 1995 (see *Nature* 379, 384; 1996).

One goal was to test suitable methods for technology foresight exercises in other areas of research. The authors agreed on a mix of methods proposed by Hariolf Grupp, vice-president of the Fraunhofer Institute for Systems and Innovation Research at Karlsruhe, and a member of the study group. Grupp's proposal combined qualitative methods, such as peer review, and quantitative—statistical methods, such as 'bibliometric mapping' of recent international publications, as well as

an analysis of a 1993 'Delphi' report on the development of science and technology.

"Many of us were rather sceptical about the use of bibliometric analysis," says Dietrich Haarer, an experimental physicist at the University of Bayreuth who has been given leave to take up the post of head of physics research at Bayer. "But it revealed interesting links; for instance the surprising frequency of cross-referencing between biophysicists and mathematicians, and between organic chemists and semiconductor researchers."

Technology foresight has been introduced in several other countries, but Germany, where scientific freedom is protected by the constitution, has tended to be sceptical. The Wissenschaftsrat had long hesitated to recommend technology foresight, fearing that it could be seen as interfering with the DFG's or MPS's funding principles. Some scientists have been concerned that foresight studies could interfere with their constitutionally guaranteed freedom to research.

But Schipanski, who promoted the strategy during her tenure as president between 1996 and 1997, denies that technology foresight is a step towards centrally directed research. "We do not intend either to replace existing principles, such as the DFG's peer-review system, or to trim basic research," she says. "Technology foresight helps to ensure common ground between scientists, politicians and the public."

Quirin Schiermeier

## Utah university finally drops out of cold-fusion patent chase

[BOSTON] A US university has stopped pursuing 'cold-fusion' patents based on the work of the chemists Stanley Pons and Martin Fleischmann. This represents "the end of a chapter" for the University of Utah, which has spent about \$500,000 in pursuing the technology, according to Richard Koehn, its vice-president for research.

The decision comes nine years after a press conference at which Pons and Fleischmann described a simple technique for producing nuclear fusion at room temperature (see *Nature* 338, 364; 1989). "After nearly a decade of work on this subject by respectable people, there has been no progress in duplicating the original claims," says Koehn. "For that reason, we decided it was not appropriate to spend any more public funds on this."

The university had been bound by its original agreements with Pons and Fleischmann to fight for patent approval in both the United States and Europe. But such endorsement of the original claims has not been forthcoming. ENECO, the company in Salt Lake City that acquired rights to both the patents and technology, relinquished its



licence last year after a fruitless and expensive attempt to win patent approval.

The university then tried to relicence the patents, but found no takers. Facing the estimated \$1 million–\$2 million in legal fees needed to pursue appeals in support of the patent applications, the university offered the licence to Pons and Fleischmann themselves, who declined. "At that point, we were finally free to step away from this technology," says Koehn.

The episode has dealt a "body blow" to the university's reputation, he says, because of the lack of technical progress in cold fusion and the way the claims were originally publicized.

But cold fusion advocates are not giving up. "The abandonment of a patent has no bearing on the science," Fleischmann said last month. "The research can proceed in a more positive way without a patent claim."

Eugene Mallove, editor-in-chief and publisher of *Infinite Energy* magazine, admits: "There is very little money in this field, but hundreds of scientists are still conducting experiments. The real action these days is in the commercial sector."

Government backing is scarce: the Japanese-sponsored research programme ended this year (see *Nature* 389, 10; 1997), and the US Department of Energy stopped funding cold fusion work in 1989.

Hal Fox, editor of the *Journal of New Energy*, thinks that cold fusion will ultimately be replaced by 'plasma-injected transmutation', a low-energy process that transforms heavy radioactive elements into stable ones.

Steve Nadis



# UK eyes social goals for next Foresight

Ehsan Masood

**Scientists met industry to discuss the technology of the future under the previous government — and it worked. The Labour government is launching a second phase, reflecting lessons learnt and its own social concerns.**

[LONDON] Britain's then Conservative government gave its backing five years ago to a broad national initiative designed to get scientists and industry thinking together about how science can help deliver the technologies, products and services of the future. It was known as Technology Foresight.

Despite the reservations of many academics at the prospect of outside interference in their research agendas, the Labour Party enthusiastically backed the idea when in opposition. Now, exactly one year into government, it has confirmed this commitment by launching a national consultation for Foresight's second phase.

The new exercise is due to start in October this year, and will last until November 2000. There are significant changes in both substance and presentation. In part, these reflect lessons learned over the past five years. But they also reflect the changed social priorities of the new government.

On substance, the widely criticized Delphi exercises of the first round, which asked respondents to fill out lengthy questionnaires predicting the technology priorities of the future, have been shelved. As for presentation, the word 'technology' has been erased; the programme will in future be known simply as 'Foresight'.

Perhaps most significant is the fact that the new Foresight exercise seeks to redress what some believe has been an imbalance between the attention previously given to its two goals — enhancing wealth creation and the quality of life respectively — by explicitly seeking to address social policy priorities.

Six themes are being provisionally suggested by the government: ageing, the future of cities, crime control, social cohesion, education and training, and sustainable development. The panels for each sector may be combined to help focus on the themes. For example, the Health and Life Sciences panels could become simply Healthcare. And the separate Energy, Natural Resources and Environment could become a single panel: Energy and Environment.

The composition of the steering committee has been radically altered, partly to reflect this shift; scientists and industrialists have been largely replaced by heads of trade bodies and social policy experts. The committee will, however, continue to be chaired by the

government's chief scientist, Sir Robert May.

By most accounts, Foresight has been a success. Sixteen panels of scientists and industrialists, covering almost the entire economy from food and drink to defence and aerospace, came up with 360 recommendations on how science might benefit each sector.

Individuals whose paths would probably have never crossed are now busy setting up collaborative ventures. Projects worth £100 million (US\$167 million) have been set up as a direct result of such interactions, with one-third of funding coming from government and the rest from industry.

"I attended all the meetings of the food-and-drink panel," says Derek Burke, then vice-chancellor of the University of East Anglia and a member of the first Foresight steering group. "Academics and people from

this industry had never talked before."

If Technology Foresight has encouraged industry to take science-based innovation more seriously, its impact on thinking in research laboratories has been even more significant. Britain's six research councils now direct virtually all their 'targeted' research funds towards Foresight priorities. Anecdotal evidence suggests that even unsolicited 'response mode' research proposals increasingly reflect Foresight themes (see box).

There has also been wide international interest. Technology forecasting is not a new phenomenon. Indeed, British Foresight was influenced by Japan, one of the world's most enthusiastic users of foresight methods. But the scale and profile of the British exercise has filtered through to many countries, particularly those with ties to Britain such as India, New Zealand and South Africa.

## Switch in emphasis

Because of its perceived success, many of the changes embodied in the new Foresight are largely in emphasis rather than methodology. For example, while scientists will continue to talk to industrialists, they will focus their collaborations more on enhancing the quality of life than merely increasing wealth creation.

But some lessons from the first exercise have been taken on board. For example, most panel members felt the Delphi exercise was

## The changing culture of British science

[LONDON] An underlying goal of Britain's Technology Foresight programme has been to change the culture of British science while protecting the basic research base.

Many believe that the programme has succeeded on both counts. Foresight has certainly altered the British scientific landscape. And the programme has also benefited science itself. Having learnt to talk to business, scientists have found an alternative source of income, useful at a time of ever-tightening government budgets.

Foresight priorities are now guiding many of the funding decisions of all six of Britain's research councils. Two in particular that have used Foresight to increase significantly their involvement with industry are the Engineering and Physical Sciences (EPSRC) and Biotechnology and Biological

Sciences Research Councils. Seventy per cent of EPSRC's research spending is now aligned with priorities identified by Foresight.

Others, such as the Natural Environment Research Council, have realigned many of their 'directed' research programmes to follow Foresight-style priorities. For example, the council recently launched a multidisciplinary Urban Regeneration and the Environment initiative to investigate how environmental and ecological research can be harnessed for urban regeneration.

The Particle Physics and Astronomy Research Council, despite a tightly defined research agenda, is also enthusiastic about the possibilities of widening the applications of particle physics and astronomy-related technologies.

Indeed, whatever their initial reservations, many

senior staff from the research councils now believe that the success of Foresight has played an important role in helping the Office of Science and Technology to protect their budgets against major cuts, and that it may also help to persuade the government to increase funding for science in its forthcoming Comprehensive Spending Review.

Each research council says that decisions on 'response mode' funding remain untouched by the Foresight culture. And many researchers remain wary of their research becoming targeted towards specific social goals.

But few deny that university scientists in Britain increasingly feel that high-quality grant applications tilted towards wealth creation and the quality of life have a better chance of being funded than those that are simply good science. **E.M.**



largely a waste of time. "It didn't tell us anything we already didn't know," says Oliver Sparrow of the Royal Institute of International Affairs in London, a member of the first Foresight steering group.

Other shortcomings of the Delphi approach have also become clear. For example, the materials panel failed to highlight nanotechnology as a focus for future research and investment.

More spectacularly, perhaps, the Information Technology panel failed to forecast what is perhaps the technological development of the decade: the Internet. The panel's report from 1995 recommends the need to exploit "the information superhighway initiative". But there is no mention of the World Wide Web and other ways in which the Internet has developed.

Another relative weakness of Technology Foresight has been its failure to reach the crucial one million-strong small to medium sized business sector, where the potential for innovation and growth tends to be higher than in larger companies.

"Foresight got to lots of big, blue-chip companies," says Paul Ormerod, programme director of the Centre for Exploitation of Science and Technology in London. "But smaller companies are usually too busy trying to stay alive to worry about the longer term." Ormerod has written a guide to foresight for small businesses, which will be launched by the government this summer.

There has also been criticism of the inability of panels to look beyond their own sector and contribute towards solving wider problems. Geoff Findlay, director of corporate affairs at the Particle Physics and Astronomy Research Council, is among those who believe that this is among Foresight's biggest failings. "It did not encourage lateral thinking," he says.

Under old Foresight, for example, ideas on a low-carbon-emitting car would have been restricted to the 'transport' panel. In contrast, in the new exercise, every relevant panel will be expected to contribute ideas to such goals. In addition to the obvious ones such as 'materials' and 'chemicals', it might include others such as 'defence and aerospace'.

Using scientific research to address social as well as economic issues ties in well with the political commitments of the Labour government. But the shift in emphasis also received overwhelming backing from respondents to a consultation on the effectiveness of the first exercise, according to the government's Office of Science and Technology (OST), which acts as the Foresight secretariat.

Many panel members, too, appear to be in favour of the idea — up to a point. "There was little point in sitting down all over again and talking about the technologies of the future," says Paul Leonard, head of research and innovation at the Chemical Industries Association, and a member of Foresight's



**New priority: how technology can help the aged is one goal of the second Foresight exercise.**

chemicals panel. "We'd done that, and needed to move on."

The benefits may extend beyond the practical. Linking Foresight to themes such as ageing and the future of cities could boost the public image of science by illustrating how it can be used to improve people's lives, suggests Burke. Such a task is important, he says, given the negative images from science following episodes such as BSE and the public response to cloning and genetically modified food.

### Boardroom appeal

Dropping the word 'technology' is another attempt to widen Foresight's constituency. Ben Martin, director of the Science Policy Research Unit at the University of Sussex, and one of the originators of the British Foresight programme, says: "There was a feeling that with technology in the title, it would be more difficult to interest the accountants and lawyers who sit in boardrooms."

Martin, one of the few survivors from the original steering group, admits he is slightly nervous about playing down the technology dimension. Those voicing similar concerns include Ian Taylor, minister of state for science and technology in the previous Conservative government, as well as many scientists and industrialists — and even some representatives of consumer groups.

Some fear that the new emphasis on social themes may turn out to be at the expense of technology and wealth creation. There is a feeling that edging out the scientists and industrialists who were critical to the first phase could risk the whole enterprise.

OST officials justify the new make-up of the steering group on the grounds that, to succeed, the next phase of Foresight needs to influence the heads of society's high-pow-

ered institutions. And that there is no better way of doing this than to co-opt them onto the steering group. But the shift worries Taylor. "It is important that we don't lose the backing of science and technology," he says.

Taylor says he has always seen the social agenda as an important part of Foresight, pointing out that there was a previous move to give Foresight a social dimension — the decision to launch the Extending Quality Life (EQUAL) initiative.

But he says too much "fuzzy thinking" should not be allowed to dominate the exercise. Harnessing science and technology for wealth creation, Taylor believes, is more important, "because without this, quality of life cannot improve".

Such concerns are shared by several panel chairmen, including in particular John Beacham of the chemicals panel, who retired last week as group research manager and chief scientist at ICI, and John Taylor, head of the Information Technology, Electronics and Communications panel and director of Hewlett Packard's European research laboratories in Bristol.

John Taylor predicts that Foresight will "evaporate if it loses its major emphasis on wealth creation". He disagrees with those who say that it is time to "move on", arguing that Foresight has "only just begun to scratch the surface of wealth creation". The first round of Foresight was too rushed and underfunded, he says. "It's important to have continuity and follow through. There's still a lot we can learn."

The consumer sector is also concerned — but for different reasons. Sheila McKechnie, director of the Consumers' Association and an observer on the food panel, says it is too soon to allow Foresight to refocus on "big, stratospheric issues".

McKechnie feels that scientists still need to be closely involved, particularly in the food and drink sector where research is critical to issues such as nutrition and food safety. "The food industry is in a state of enormous upheaval," she says. "We need research to tell us how to eliminate scrapie and salmonella, and whether using antibiotic markers in grain is a good or a bad thing."

The level of scientific involvement is clearly a controversial issue, cutting to the heart of the new Foresight exercise. But Diane Coyle, economics editor of Britain's daily national newspaper *The Independent* and newly appointed member of the Foresight steering committee, thinks that having fewer scientists on the steering group is probably a good thing.

"Scientists tend to be quite deterministic about things. They don't always understand the links between science and society," she says. Given that the purpose of Foresight is to make an impact on business, the economy and society, Coyle feels it important to involve individuals from those sectors directly. □

## Engineered foods not 'organic', says US agriculture department

[WASHINGTON] The US Department of Agriculture has decided to exclude genetically engineered and irradiated foods from the designation 'organic', it was reported last week.

The department had previously proposed a definition of 'organic' that did not exclude these foods. But during a four-month comment period that ended last week, it received some 150,000 communications from the public, most vociferously opposing their inclusion.

Even the agricultural giant Monsanto argued against including genetically engineered foods. A final definition is expected from the department by the end of the year.

Meanwhile, Britain's Vegetarian Society, which has been criticized for allowing its logo indicating approval to be used on foods containing genetically modified soya, is coming under pressure from the producers of 'natural' foods to modify its stance. One critic argued last week that the society had been 'duped' into approving such soya, which can be used as an alternative to meat.

## Euro parliament holds out for more Framework cash

[MUNICH] The European Parliament is likely to reject the European council of research ministers' proposal for financing the fifth Framework programme of research (FP5), which should be adopted in the autumn and run for five years.

Council and parliament disagree strongly about the level of financing of FP5 which the council, following a meeting in February, wants restricted to ECU14 billion (US\$15.6 billion). After its first reading, parliament had asked for ECU16.7 billion.

Parliament's research committee is due to discuss proposed amendments to the council's proposal on 20 May, in preparation for a second reading in parliament at the beginning of June. It is likely to hold out for a minimum of ECU15 billion.

## Competition to launch French start-ups

[PARIS] A competition to help young scientists to create high-technology start-up companies has been started by the Rhône Poulenc Foundation, a charity set up by the company Rhône Poulenc under the auspices of the Institut de France. Ten so-called 'springboard' projects will be selected annually by a jury chaired by Pierre-Gilles de Gennes, the 1991 Nobel prizewinner in physics.

Each winner will receive funding of FF2 million (US\$335,000), and free assistance for a year from a senior consultant, who will provide business expertise and contacts. Winners will also be given training in the creation of enterprises at the University of Paris-Dauphine, including courses on financial management, the financing of innovation and market research.

## Peking University's 100th birthday draws crowds

[BEIJING] About 100,000 alumni of Peking University were expected to turn up this week for the university's centennial celebrations on 4 May. University officials said that more than 10,000 were expected from overseas, and some jumbo jets were reportedly chartered from US cities to fly them in.

There has been frenzied activity on the campus recently to repave the roads and repaint the buildings for the celebrations, which will provide a vivid demonstration of the power of the network that links Chinese academics around the globe.

## Earthquake scientists win Crafoord Prize

[WASHINGTON] Two earthquake specialists working in the United States have won this year's Crafoord Prize for "their fundamental contributions to our knowledge of the structures and processes in the interior of the Earth". Geophysicist Don L. Anderson, of the California Institute of Technology, and geologist Adam M. Dziewonski, of Harvard University, are to share the \$500,000 prize, which is awarded by the Royal Swedish Academy of Sciences for disciplines not covered by the Nobel prizes.

Anderson has shown how changes in the composition of the Earth's mantle can cause tensions in the planet's crust that lead to earthquakes, and Dziewonski's work on seismic waves has led to the development of computer tomography to measure and analyse global seismic data.

## Congress funds research to speed up Internet

[WASHINGTON] Funding of \$57 million for university research into faster Internet links will become available at the National Science Foundation (NSF) as part of a \$6 billion supplemental funding bill signed into law by President Bill Clinton last Friday (1 May).

A provision in the bill releases \$57 million, which the NSF has acquired from licensing payments received whenever someone registers a domain name on the Internet. Early last month, a US District Court judge called into question the legality of these payments, and ruled that the money could

not be spent until Congress said that the funds had been collected appropriately. The bill does that, according to NSF officials.

## Four share Australian award for genetics

[SYDNEY] The 1998 government-funded Australia Prize for molecular genetics, worth A\$300,000 (US\$195,000), has been shared by four scientists. Suzanne Cory was recognized for research, done mainly with her husband, Jerry Adams, at the Walter and Eliza Hall Institute in Melbourne, that includes the discovery of the genetic mutation leading to Burkitt's lymphoma.

The second Australian winner was Grant Sutherland, working at the Adelaide Children's Hospital, who first mapped fragile sites on the X chromosome. Elizabeth Blackburn of the University of California, San Francisco, an Australian by birth, was rewarded for her work on the discovery of telomerase, and Sir Alec Jeffreys, from the University of Leicester in the United Kingdom, was honoured for the discovery in 1984 of DNA fingerprinting.

## Rüttgers honoured for promotion of biotech

[MUNICH] Jürgen Rüttgers, the German science minister, was last week named as Best European Life Science Entrepreneur of the Year by the European Life Sciences Partnering Foundation, a body set up in 1992 by the venture capital company Atlas Venture and the business consultancy Ernst and Young.

Rüttgers was awarded the annual prize for his "exceptional contribution to the development of young European biotechnology companies". To promote biotechnology start-ups, Rüttgers set up the so-called BioRegio competition two years ago, and loosened the restrictive German legislation on genetic engineering.

## World Bank signs deal to protect Brazil's forests

[LONDON] The World Bank and the World Wide Fund for Nature (WWF) last week announced a joint agreement with the government of Brazil to protect 62 million acres of the Amazonian rainforest against development and slash-and-burn agriculture.

The estimated cost of between US\$100 million and US\$350 million will come from a combination of soft loans and cash from the Global Environment Fund, set up after the 1992 Earth Summit. The agreement represents the first result of a partnership between WWF and the bank announced last year. It is hoped that 125 million acres of forest will be protected worldwide by 2005.

# It's Jakob's disease, not Creutzfeldt's

Sir — Creutzfeldt–Jakob disease (CJD) is a misnomer. In 1921, the Hamburg neuropathologist Alfons Maria Jakob (1884–1931) published three papers, “Über eigenartige Erkrankungen des Zentralnervensystems mit bemerkenswertem anatomischen Befunde” (“On peculiar illnesses of the central nervous system with remarkable anatomical findings”)<sup>1–3</sup>.

In these and in a paper of 1923 (ref. 4), he described the cases of two men and three women aged between 34 and 51 who progressively developed disturbances of motor functions, speech and emotion, with obvious personality changes and loss of memory; eventually they were unable to move, stand or speak, and died in dementia between a few weeks and a year after the start of more serious symptoms.

In 1920, the Breslau neurologist Hans Gerhard Creutzfeldt (1885–1964) had described “a peculiar focal illness of the central nervous system” of the patient Berta E.<sup>5</sup> Jakob writes in his second paper on these new diseases that he came to the conclusion that Creutzfeldt's case refers to a “nosologically very closely connected if not identical affection” (“*nosologisch sehr nahestehende, wenn nicht wesensgleiche*

*Affektion*”) to his own cases<sup>6</sup>. Thus the designation “Creutzfeldt–Jakob–Krankheit” was logical. It had first been used in 1922 (ref. 7).

Medicine today considers CJD as a rare, progressive and always deadly transmissible spongiform encephalopathy (TSE), and also as a prion disease. Careful perusal of the case reported by Creutzfeldt on the basis of the clinical and pathological findings and his illustrations showed that his case does not belong to the TSE group.

E. E. Manuelidis reports: “Dr Creutzfeldt after the Second World War told me that his case did not bear any resemblance to the cases described by Jakob”<sup>8</sup>. Strictly speaking, therefore, the sickness should be called Jakob's disease. It was of course Jakob's own fault that his name is connected erroneously with Creutzfeldt.

But, even of Jakob's five cases, only two are what we understand today as CJD. C. L. Masters was able to re-examine the original slide preparations of Jakob's cases still preserved in the archives of the department of neurology at the University of Hamburg, and he found that only the third and fifth of Jakob's cases fall within the present diagnostic criteria. Five of six

subsequent cases published from Jakob's laboratory, however, under his supervision, showed the typical changes of spongiform encephalopathy<sup>9</sup>.

Because of bovine spongiform encephalopathy, Creutzfeldt–Jakob disease has become almost a household name. It seems impossible, therefore, to drop the Creutzfeldt part of this term because everybody would think that Jakob's disease is something different. But it would be possible to do Jakob historical justice by changing the order of the two names so that his comes first, and by always using the term Jakob–Creutzfeldt disease as is often done in German medical literature.

**Friedrich Katscher**

*Mariahilfer Str. 133,  
A-1150 Vienna,  
Austria*

1. Jakob, A. D. Z. Nervenheilk. 70, 132–146 (1921a).

2. Jakob, A. Z. ges. Neurol. Psychiatr. 64, 147–228 (1921b).

3. Jakob, A. Med. Klin. 17, 382–386 (1921c).

4. Jakob, A. Monogr. Gesamtgeb. Neurol. Psychiatr. 37, 215–345 (1923).

5. Creutzfeldt, H. G. Z. ges. Neurol. Psychiatr. 57, 1–18 (1920).

6. Jakob, A. Z. ges. Neurol. Psychiatr. 64, 174–175 (1921).

7. Spielmeyer, W. Klin. Wochenschr. 1, 1817–1819 (1922).

8. Manuelidis, E. E. J. Neuropathol. 44, 1–17 (1985).

9. Masters, C. L. & Gajdusek, D. C. in *Recent advances in Neuropathology* Vol. 2 (eds Smith, W. T. & Cavanagh, J. B.) 139–163 (Churchill Livingstone, Edinburgh, 1982).

## Patents paved the way

Sir — As the original patentee of the nuclear radiometer concept, I am impressed with Daedalus's imaginative applications of the idea (*Nature* 392, 337 & 443; 1998).

The basic patent (Self-Propelled Nuclear Radiometer, US Patent No. 3,110,811) almost 40 years ago employed precisely the principle that Daedalus describes — to create thrust for small motors or in outer space. Five years earlier, that principle was employed in another patent (Nuclear Radiometer for Neutron Flux Measurement, US Patent No. 2,951,942), using a neutron absorber such as fissionable materials or boron-10, to create particles to heat one side of radiometer panes to measure neutron flux.

**Arnold Kramish**

*2065 Wethersfield Court,  
Reston, Virginia 20191, USA*

## Training undervalued

Sir — The German labour market is in its deepest crisis of the post-war era. One of the severest shortcomings in Germany is a lack of willingness to take young people

into apprenticeship. More than 200,000 such positions are likely to be needed each year. At the same time, the effectiveness of research activities inside and outside universities is lagging behind that of US and British institutions. The possible causes and potential remedies are widely discussed in the media. I should like to suggest a way in which the German science community might help to solve Germany's labour problem and at the same time increase its productivity.

Technical personnel in Anglo-Saxon countries usually have a university education, often a bachelor's degree. As a consequence, their education is up to date with the most modern technologies and knowledge. In Germany, by contrast, equivalent workers are educated either within companies that employ largely traditional methods or in schools whose curricula are years behind the state of the art. Young workers newly employed by research institutions and technology companies have therefore to be entirely retrained to be able to perform as desired. Because such retraining is usually unsystematic, the theoretical background of technical personnel in Germany is often insufficient to allow their creative participation in research.

Large research institutions, even those entirely financed through the government, only very rarely offer apprenticeships. Academics are extremely reluctant to educate technical personnel, considering such activity a burden. But the blame does not lie only with academics. Government and large research institutions do little to convince their employees of the need to educate young people or otherwise promote activities along these lines. For example, very few positions would have to be created in big institutes that coordinate teaching. Apprentices could rotate through different laboratories and acquire profound and versatile skills.

This would not only not be a burden, but within a few years would result in much better educated staff and, in consequence, higher research productivity. Thousands of positions could be created by this means, and each graduate would certainly be sufficiently qualified to compensate for the trouble caused to the scientists involved in training.

**Alexander Olek**

*Max-Planck-Institut für Molekulare Genetik,  
Innestr. 73,  
D-14195 Berlin,  
Germany  
e-mail: olek@mpimg-berlin-dahlem.mpg.de*



# The burst, the burster and its lair

Ralph Wijers

**An analysis of the light from a  $\gamma$ -ray burst shows that it occurred in a very distant galaxy. This confirms that bursts are by far the most powerful bombs in nature's arsenal, and only the most copious known energy source can explain them.**

In 1969, two satellites were launched to watch for covert nuclear weapons tests. Instead, they detected occasional flashes of  $\gamma$ -rays, coming from apparently random directions in space. Since then, the origin of these  $\gamma$ -ray bursts (GRBs) has been among the greatest problems in astrophysics. Because there was so little information to go on — particularly because there was no indication of the distance to bursts — theories were unconstrained. Ideas ranged from quakes on the surfaces of neutron stars in our Galaxy, to merging pairs of neutron stars in distant galaxies, to tiny exploding black holes at the edge of our Solar System.

But  $\gamma$ -ray-burst research has been revolutionized since the first discovery of an optical afterglow<sup>1</sup>, last year. The BeppoSAX satellite, by rapidly determining the location of several bursts, has allowed follow-up observations to see the remnant of burst activity at longer wavelengths than those of  $\gamma$ -rays. Among other things, this finally settled the distance scale: the optical spectrum of GRB970228 had a redshift<sup>2</sup> of at least 0.835, showing that it occurred in a galaxy halfway across the Universe. That means the explosion must have been enormously energetic.

But a burst seen on 14 December 1997 has told us even more. Three papers in this issue describe how GRB971214 occurred at an enormous redshift of 3.42 (when the Universe was only one-seventh of its present age) in an actively star-forming galaxy<sup>3</sup>, and was somewhat obscured by dust<sup>4,5</sup>.

Halpern *et al.*, on page 41 of this issue<sup>4</sup>, describe how the optical-to-infrared light

from GRB971214 appears to have been dimmed and reddened by dust in its host galaxy, which might suggest that the burst happened in a gas-rich environment. The spectrum also tells us that the host galaxy was producing stars at a high rate. Both the high redshift and the location in a star-forming galaxy support the idea that GRBs are related to dying stars<sup>6,7</sup>. If so, we might find bursts that mark the deaths of the very first stars that ever formed, and thus learn about a period in the history of our Universe that we cannot see in any other way.

In order for  $\gamma$ -ray photons to escape from the burst (rather than be absorbed by the explosion), the radiating material must be expanding at nearly the speed of light<sup>8</sup>. Radio observations of the afterglow of GRB970508 confirmed this relativistic expansion directly<sup>9</sup>, and the optical afterglows of other bursts have the right properties to be coming from the collision of these relativistic ejecta with the surrounding gas<sup>10</sup>. On page 43 of this issue<sup>3</sup>, Ramaprakash *et al.* use a change of slope in the optical spectrum to show that the afterglow had an energy of at least  $10^{51}$  erg — although absorption due to dust<sup>11</sup> could have contaminated this result. But another lower limit to the energy is set by Kulkarni *et al.*, on page 35 of this issue<sup>3</sup>, who calculate from the high redshift that the burst emitted  $3 \times 10^{53}$  erg in  $\gamma$ -rays alone. This is around 20% of the rest-mass energy of the Sun, or 50 times as much as the Sun will radiate in its lifetime.

So what could provide this huge energy in the very small volume indicated by the rapid variability of GRBs? The gravitational

collapse of about a solar mass of matter into a neutron star or black hole would naturally satisfy both requirements. The oldest such burster model is the merger of two neutron stars, which we expect to happen a few times in a million years in our Galaxy. But it had been thought that no more than about  $10^{51}$  erg of energy could be generated in this manner.

Another leading model is that GRBs occur when the core of a very massive star collapses directly into a black hole, leaving a dense torus of material briefly orbiting. Such a 'failed supernova'<sup>12</sup> or 'hypernova'<sup>13</sup> could generate plenty of energy, from the accretion of the torus onto the black hole and from the extraction of spin energy from the black hole. But as the merger of two neutron stars will also form a black hole surrounded by a torus, the same larger energy sources are available there<sup>14</sup>. So it would be premature to write the obituary of neutron star merger models. Moreover, all of these forms of gravitational collapse produce rapidly spinning objects, which probably spew out their energy in two narrow beams, along the axis of rotation, rather than in all directions. Consequently, by assuming that GRB emission is spherical, we may be overestimating their energy output by a factor of 10 to 100.

An important difference between the models is the location of bursts. It is striking that all optical afterglows for which we have precise (sub-arcsecond) locations to date (970228, 970508, 971214, 980326, 980329) are right on top of a host galaxy. This does argue against merging neutron stars (which we expect to have travelled many thousands of light years before they collide and cause a burst) and in favour of models such as hypernovae, which are expected to happen within star-forming regions<sup>13</sup>.

It is clear that constraints on physical models of GRBs are beginning to emerge at last, but firm conclusions may await an answer to the question of whether emission is beamed. That can be tested observationally: if  $\gamma$ -rays are strongly beamed, then as the optical afterglow is less beamed there should be many 'homeless' optical afterglows (ones not preceded by a GRB). A search for these would be more difficult than the current searches for high-redshift supernovae, but feasible with modern instruments.

The measurement of afterglow spectral breaks is a potentially powerful method for deriving constraints on physical parameters such as the total explosion energy and the density of the surrounding medium<sup>5,11</sup>. As spectral breaks will appear in optical wavelengths for only a day, swift detection and multicolour photometry in the first hours will be crucial. If you have a one-metre telescope, you could discover the next afterglow or contribute vital data, especially if your longitude is not covered by large observatories in your hemisphere. Look on

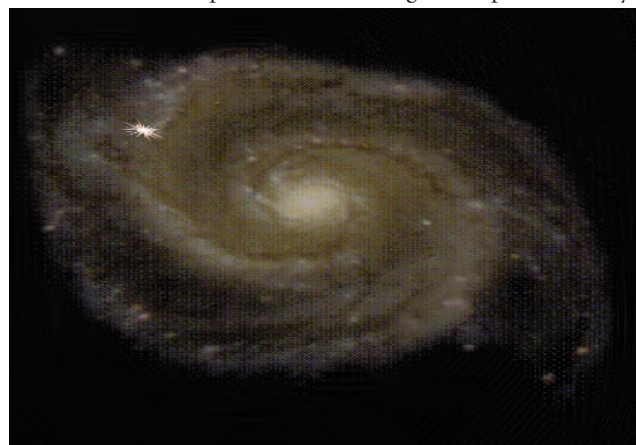


Figure 1 Artist's impression of a  $\gamma$ -ray burst's optical afterglow, shining out from a star-forming region in a spiral galaxy. The latest observations hint that bursts originate in such dusty regions in galaxies, and favour the idea that they result from collapsing stellar cores.

[http://gcn.gsfc.nasa.gov/gcn/gcn\\_main.html](http://gcn.gsfc.nasa.gov/gcn/gcn_main.html) to find out how to be alerted instantly when a GRB is located accurately.

As we find fainter bursts, we may be investigating the early Universe as well as the ultrarelativistic fireworks themselves. But because all the visible light from very distant bursts will be absorbed by gas along the line of sight (as the bluest light from GRB971214 already is), we need to look for them in the infrared. The weaker signals from these bursts will also make their locations less accurately measurable, so the hunt for these distant echoes of the first star formation will be made by the new generation of wide-field infrared cameras that are just coming on line.

All the issues that Bohdan Paczyński and I raised last year<sup>15</sup> have already been resolved: all the GRBs detected by BeppoSAX come from beyond our Galaxy; they all appear to lie within faint galaxies (some definitely outside galactic nuclei); most have X-ray afterglows and nearly half have optical afterglows. Some cases of absent optical

afterglows when X-rays were seen could be due to strong dust absorption<sup>16</sup>.

It might seem optimistic to think that all the issues raised here will once again be solved within a year, but the pace of discovery is still very rapid. □

Ralph Wijers is at the Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK. e-mail: ramjw@ast.cam.ac.uk

1. van Paradijs, J. *et al.* *Nature* **386**, 686–689 (1997).
2. Metzger, M. *et al.* *Nature* **387**, 878–880 (1997).
3. Kulkarni, S. R. *et al.* *Nature* **393**, 35–39 (1998).
4. Halpern, J. P., Thorstensen, J. R., Helfand, D. J. & Costa, E. *Nature* **393**, 41–43 (1998).
5. Ramaprakash, A. N. *et al.* *Nature* **393**, 43–46 (1998).
6. Totani, T. *Astrophys. J.* **486**, L71–L74 (1997).
7. Wijers, R. A. M. J., Bloom, J. S., Bagla, J. S. & Natarajan, P. *Mon. Not. R. Astron. Soc.* **294**, L13–L17 (1998).
8. Zdziarski, A. A., Svensson, R. & Paczyński, B. *Astrophys. J.* **366**, 343–352 (1991).
9. Frail, D. A., Kulkarni, S. R., Nicastro, L., Feroci, M. & Taylor, G. B. *Nature* **389**, 261–263 (1997).
10. Mészáros, P. I. & Rees, M. J. *Astrophys. J.* **476**, 232–237 (1997).
11. Galama, T. J. *et al.* *Astrophys. J. Lett.* (in the press).
12. Woosley, S. E. *Astrophys. J.* **405**, 273–277 (1993).
13. Paczyński, B. *Astrophys. J.* **494**, L45–L48 (1998).
14. Mészáros, P. I. & Rees, M. J. *Astrophys. J.* **482**, L29–L32 (1997).
15. Paczyński, B. & Wijers, R. A. M. J. *Nature* **386**, 650–651 (1997).
16. Groot, P. J. *et al.* *Astrophys. J.* **493**, L27–L30 (1998).

## Membrane fusion

# SNAREs line up in new environment

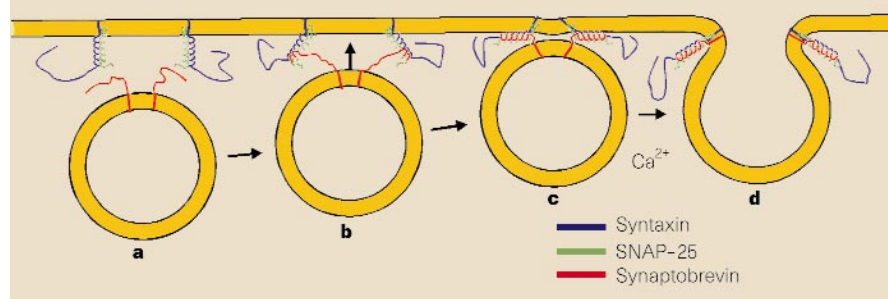
Reinhard Jahn and Phyllis I. Hanson

Many cellular processes — ranging from growth to neuronal communication — rely on fusion of intracellular membranes. This requires a highly conserved set of membrane proteins known as SNAREs. Relatives of these proteins are involved in almost every type of intracellular fusion, and probably form the core of the fusion machinery<sup>1,2</sup>. But how exactly do SNAREs mediate fusion, and how do the other proteins that are known to interact with them fit into the picture<sup>1,2</sup>? Functional<sup>3,4</sup> and structural<sup>5,6</sup> studies, together with a reconstitution of SNAREs into lipid-bilayer vesicles reported by Weber *et al.*<sup>7</sup> in *Cell*, have begun to shed some light on the inner workings of SNARE proteins and their effects on membranes.

The best-characterized SNAREs are involved in neuronal exocytosis. They include the synaptic-vesicle protein synaptobrevin, and the plasmalemma-associated proteins syntaxin and SNAP (synaptosome-associated protein)-25. Treatment with botulinum and tetanus neurotoxins — which selectively digest these three SNAREs — blocks exocytosis<sup>8</sup>, providing compelling evidence for their central role in this process. Syntaxin and synaptobrevin are anchored in the lipid bilayer by transmembrane regions found at their carboxy termini, whereas SNAP-25 is anchored by palmitoyl side-chains. The three proteins assemble spontaneously into a ternary complex, which can be disrupted by the ATPase NSF (*N*-ethyl-

maleimide-sensitive fusion protein), together with SNAPs (soluble NSF-attachment proteins). Because SNAPs mediate binding of NSF to this ternary complex, its components are collectively called SNAREs (SNAP receptors)<sup>1</sup>.

Membrane-fusion processes throughout the cell seem to require the cyclic assembly and disassembly of such ternary complexes<sup>1</sup>.



**Figure 1** Emerging model for protein-mediated membrane fusion in neuronal exocytosis. This scheme is based on functional<sup>3,4,7,15</sup> and structural<sup>5,6</sup> work. Weber *et al.*<sup>7</sup> have reconstituted the proteins involved in membrane fusion into artificial membranes, and their results support this scheme. a, Syntaxin and SNAP-25 form a complex on the plasma membrane. b, After initial contact, unfolded synaptobrevin on the vesicle forms a ternary complex with the syntaxin/SNAP-25 complexes, generating a tight bundle of  $\alpha$ -helices connected at one end to the two membranes. c, Because the assembling helices comprise membrane-proximal regions of syntaxin and synaptobrevin, assembly of the ternary complex forces their membrane anchors into close apposition, drawing the underlying membranes close together. For constitutively fusing membranes, such complex assembly may be enough to induce fusion. In neurons, fusion is probably triggered by  $\text{Ca}^{2+}$ -induced interactions involving synaptotagmin, the neuronal  $\text{Ca}^{2+}$  receptor<sup>2</sup>. d, After fusion, stable ternary complexes are anchored in a single fused membrane. These are disassembled by the ATPase NSF, to allow synaptobrevin to be recycled into vesicles, and to free syntaxin and SNAP-25 for new rounds of membrane fusion with incoming synaptic vesicles.

How does this relate to membrane fusion? Membranes that are destined to fuse with one another contain complementary sets of SNAREs, designated v (for vesicle) and t (for target) SNAREs. Originally, membranes were thought to be joined by the assembly of complexes containing v- and t-SNAREs in an antiparallel arrangement, with carboxy-terminal transmembrane anchors on opposite ends. These docking complexes were believed to be then rearranged by NSF, getting them out of the way while promoting the fusion reaction<sup>1,9</sup>. But several new lines of work have changed this view.

First, Wickner and colleagues<sup>3,4</sup> have characterized the molecular requirements for fusion between vacuole-precursor vesicles from bakers' yeast, *Saccharomyces cerevisiae*, using genetics together with an *in vitro* assay. They confirmed that complementary SNAREs need to be present in the partner membranes<sup>3</sup>, but they also found that NSF is not involved in fusion<sup>4</sup>. Second, electron microscopy and fluorescence resonance energy transfer experiments have shown that the proteins in the ternary v-SNARE complexes are aligned in parallel, with their transmembrane anchors extending from one end of a rod-shaped particle<sup>5,10</sup>. Finally, these complexes have been found to be extraordinarily stable, with a more folded and  $\alpha$ -helical structure than any of the individual proteins. So, assembly of the ternary complexes is associated with large conformational changes and with the release of energy<sup>6</sup>.

These findings led to the proposal that the formation — rather than disassembly — of ternary complexes drives membrane fusion. Assembly could proceed from the distal tips of the individual SNAREs towards

their respective membrane anchors, thereby forcing the membranes close together and overcoming the energy barrier that normally prevents fusion<sup>5,11</sup> (Fig. 1). Interestingly, a similar mechanism has been suggested to explain how enveloped viruses infect cells<sup>12</sup>. These viruses inject their genetic material using specialized membrane proteins that mediate fusion between the viral membrane and that of a target cell. After appropriate activation, viral fusion proteins change conformation and can insert a newly exposed amphiphilic fusion peptide into the target membrane<sup>13</sup>. Subsequent conformational changes seem to allow the now doubly anchored protein to fold back on itself, pulling the fusing membranes together<sup>12,13</sup>.

The SNARE complex has, until now, mainly been studied using proteins in solution, without the membranes that they ultimately control. Weber *et al.*<sup>7</sup> have taken the step of successfully reconstituting SNAREs into artificial membranes, allowing these proteins to be studied in a defined environment. With synaptobrevin in one membrane and syntaxin and SNAP-25 preassembled in another, they found that these proteins could form ternary complexes and induce interactions between the membranes. The interactions were observed as slow mixing of the membrane bilayers, suggesting that the vesicles had spontaneously fused with one another. Thus, the idea that SNAREs can interconnect — and possibly fuse — membranes has been realized in a purified system, complementing results obtained using the complex biological membranes of yeast vacuole-precursor vesicles<sup>3</sup>.

These models of SNARE function are almost certainly oversimplified. But they do account for most of the available data. The use of artificial membranes, such as those described by Weber *et al.*, will open the way to understanding the molecular mechanisms that underlie fusion of biological membranes, while distinguishing them from nonspecific fusion caused by membrane-perturbing molecules such as annexins and amphiphilic peptides<sup>14</sup>. Much will be learned by understanding how (and how many) ternary complexes are arranged at sites of fusion, and how regulatory proteins such as munc-18/Sec1, Rab/ypt and the calcium sensor synaptotagmin interact with and affect such complexes. These approaches should help us to work out how biological fusion acquires its characteristic efficiency, speed and regulation. □

Reinhard Jahn is in the Department of Neurobiology, Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg 11, D-37077 Göttingen, Germany.  
e-mail: rjahn@gwdg.de

Phyllis I. Hanson is in the Department of Cell Biology and Physiology, Washington University School of Medicine, St Louis, Missouri 63110, USA.  
e-mail: phanson@cellbio.wustl.edu

1. Rothman, J. E. *Nature* **372**, 55–73 (1994).
2. Südhof, T. C. *Nature* **375**, 645–653 (1995).
3. Nichols, B. J. *et al.* *Nature* **387**, 199–202 (1997).
4. Mayer, A., Wickner, W. & Haas, A. *Cell* **85**, 83–94 (1996).
5. Hanson, P. I. *et al.* *Cell* **90**, 523–535 (1997).
6. Fasshauer, D. *et al.* *J. Biol. Chem.* **272**, 28036–28041 (1997).
7. Weber, T. *et al.* *Cell* **92**, 759–772 (1998).
8. Montecucco, C. & Schiavo, G. Q. *Rev. Biophys.* **28**, 423–472 (1995).
9. Söllner, T. *et al.* *Cell* **75**, 409–418 (1993).
10. Lin, R. C. & Scheller, R. H. *Neuron* **19**, 1087–1094 (1997).
11. Hanson, P. I., Heuser, J. E. & Jahn, R. *Curr. Opin. Neurobiol.* **7**, 310–315 (1997).
12. Weissenhorn, W. *et al.* *Nature* **387**, 426–430 (1997).
13. Hernandez, L. D., Hoffman, L. R., Wolfsberg, T. G. & White, J. M. *Annu. Rev. Cell Dev. Biol.* **12**, 627–661 (1996).
14. Stegmann, T., Doms, R. W. & Helenius, A. *Annu. Rev. Biophys. Chem.* **18**, 187–211 (1989).
15. Pelham, H. R. B. *Eur. J. Cell Biol.* **74**, 311–314 (1997).

## Nanotechnology

# Carbon-based electronics

Paul L. McEuen

Microelectronics is probably the technology that has most changed the twentieth century, and silicon is the atom at its heart. Although electronics, and hence silicon, will undoubtedly remain of central importance as we move into the next millennium, many would argue that biotechnology will be the most revolutionary technology of the early twenty-first century. If this is so, then silicon's upstairs neighbour in the periodic table, carbon, will take over as the lead atom. It could even do so in silicon's traditional patch, electronics, and an early sign of that promise is the paper by Tans *et al.* on page 49 of this issue<sup>1</sup>, which describes a transistor made from a single large molecule — a carbon nanotube.

First, what makes carbon's established competitor so useful? As a group-IV element, silicon arranges itself into four-fold-coordinated  $sp^3$ -bonded crystals. Charge carriers can be introduced into the material in a variety of ways to make electronic devices. Small amounts of dopant atoms create either free electrons (for an n-type semiconductor) or free holes (p-type), and differently doped regions can be combined to yield devices such as p-n junction diodes and bipolar transistors. Alternatively, a metallic gate can be used to electrostatically attract electrons or holes to the interface of silicon with  $SiO_2$ , creating a metal-oxide field-effect transistor (MOSFET). These silicon-based devices can be manufactured with near perfection

and interconnected in extraordinarily precise ways to create information storage and computing machines.

Despite having the same number of valence electrons as silicon, carbon is a more flexible atom when it comes to bonding. This flexibility is behind its usefulness in biology as the backbone of biomolecules. It also means that there are two crystalline forms of solid carbon:  $sp^3$ -bonded diamond and  $sp^2$ -bonded graphite. At first glance, neither of these seems to be particularly well-suited for electronics applications. In its diamond form, carbon is an insulator, with a gap that is too large for ordinary electronics (although it has been useful in high-temperature applications). In its  $sp^2$ -bonded form, graphite, carbon is a semimetal. It forms two-dimensional sheets with conducting states existing only along certain directions in momentum space. Although these semimetallic sheets are fascinating from a physics point of view and can be used to make simple resistive electronic elements, they are not very useful for advanced electronics.

The key development in carbon electronics came with the realization that these graphite sheets can be rolled up into tubes, which quantizes the momentum of the electrons moving around the circumference of the tube. That results in tubes that are either one-dimensional metals or semiconductors, depending on how the

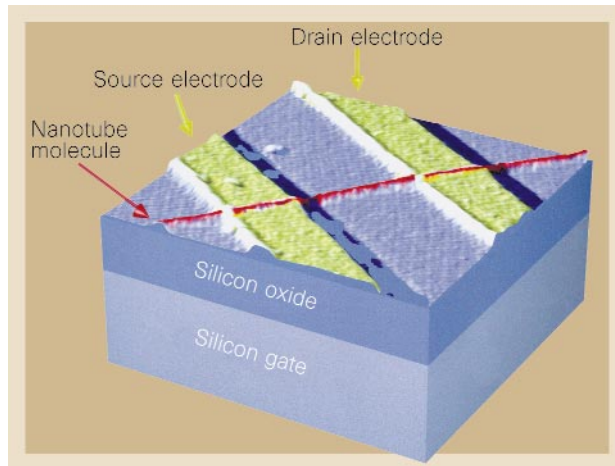


Figure 1 A nanotube transistor. A semiconducting carbon nanotube about one nanometre in diameter bridges source and drain electrodes, and a gating voltage applied to the silicon substrate induces carriers onto the nanotube to turn the transistor on.



allowed momentum quantization values compare with the preferred directions for conduction. In other words, the electronic properties of tubes are controlled by their structural details<sup>2</sup>. This remarkable prediction has recently been verified using a scanning tunnelling microscope to image the atomic structure of tubes and probe their electronic properties<sup>3,4</sup>.

By attaching leads to nanotubes, electronic devices are created. Metallic tubes have already been made into single-electron transistors that operate at low temperatures<sup>5,6</sup>, but Tans *et al.*<sup>1</sup> now provide the first report of a nanotube transistor based on semiconducting tubes (Fig. 1). A nanotube lies across two metallic contacts fabricated on top of a layer of SiO<sub>2</sub>, and a voltage is applied to the conducting substrate to move carriers onto the tube. The authors find that the tube can be 'turned on' by applying a negative bias to the substrate, which induces holes on the initially non-conducting tube. This device is thus analogous to a p-type MOSFET, with the nanotube replacing silicon as the material that hosts charge carriers. The resistance of the device can be changed by many orders of magnitude, and it operates at room temperature — a property that has eluded most other nanoscale devices.

This experiment is a first step in developing electronic devices based on semiconducting nanotubes. Other advances are likely to follow soon. Doping of the nanotubes is possible, and has been demonstrated in ensembles of tubes<sup>7</sup> and more recently in single ropes (M. Bockrath, personal communication). Other devices, such as p–n junction diodes and bipolar transistors, are likely to be realized soon. And more exotic devices are possible, owing to the unusual properties of nanotubes. For example, a single defect can change the structure of a tube from the metallic to the semiconducting variety, creating a metal–semiconductor junction composed entirely of carbon atoms<sup>8–11</sup>. Such devices would be free from problems that occur in conventional devices made from different materials, such as interdiffusion. They would also, of course, be much smaller, and therefore much faster.

The success of carbon-based electronics will depend on how rapidly the techniques for fabricating, doping and manipulating nanotubes develop. As yet these techniques are crude, and dreams of a controlled technology seem fanciful; but many groups are working hard on these issues, and their progress is impressive. By combining techniques from engineering, chemistry and biology, researchers are learning how to grow, cut, sort and chemically modify nanotubes in new and exciting ways. It is difficult to imagine carbon-based electronics competing head-on with silicon in the near

future, but there may be niche applications to which it is particularly well-suited — and if carbon is indeed the atom of the twenty-first century, then history is on its side. □

Paul L. McEuen is at the University of California and Lawrence Berkeley National Laboratory, 366 LeConte Hall, Berkeley, California 94720, USA.  
e-mail: mceuen@physics.berkeley.edu

1. Tans, S. J., Verschuereen, A. R. M. & Dekker, C. *Nature* **393**, 49–52 (1998).

2. Dresselhaus, M. S. *Nature* **391**, 19–20 (1998).
3. Wildoer, J. W. G. *et al.* *Nature* **391**, 59–62 (1998).
4. Odom, T. W. *et al.* *Nature* **391**, 62–64 (1998).
5. Tans, S. J. *et al.* *Nature* **386**, 474 (1997).
6. Bockrath, M. *et al.* *Science* **275**, 1922–1924 (1997).
7. Lee, R. S., Kim, H. J., Fischer, J. E., Thess, A. & Smalley, R. E. *Nature* **388**, 255–257 (1997).
8. Chico, L. *et al.* *Phys. Rev. Lett.* **76**, 971–974 (1996).
9. Saito, R. *et al.* *Phys. Rev. B* **53**, 2044–2050 (1996).
10. Charlier, J. *et al.* *Phys. Rev. B* **53**, 11108–11113 (1996).
11. Collins, P. G. *et al.* *Science* **278**, 100–103 (1997).

## Conservation biology

# What killed the monk seals?

John Harwood

In May and June 1997 the bodies of over 100 Mediterranean monk seals (*Monachus monachus*) were found along the Cap Blanc peninsula. This short stretch of coast, which spans the border between Mauritania and the former Western Sahara, is home to the largest surviving colony of these critically endangered creatures. The species' highly fragmented distribution extends as far east as the Black Sea, but the only other population of any size is over 4,000 km from Cap Blanc, in the eastern Mediterranean. Few ecologists were surprised when this mass mortality was attributed to infection with a previously undescribed morbillivirus<sup>1</sup>, for viruses in this family have caused similar catastrophes in other species of marine mammal. The seals in the affected area are also virtually confined to two caves only about 1 km apart, so providing ideal conditions for the transmission of infection.

On page 28 of this issue, however, Hernández *et al.*<sup>2</sup> suggest that the seals actually died because they had eaten fish contaminated with phycotoxins, which are produced during blooms of certain dinoflagellate algae. The same toxins cause paralytic shellfish poisoning in humans. Neither of the algal species implicated in the die-off had been recorded on this coast

before 1994, when one of the species was responsible for the death of four people.

Identifying the cause of mass mortalities in wildlife populations is always tricky, and could have been particularly difficult in this incident (the political status of the Western Sahara has been undefined for several decades, making access difficult, and the whole area is littered with landmines). However, the Cap Blanc seals have been studied by a team from the University of Barcelona since 1993 (ref. 3), and many individuals can be recognized by their unique markings. Also, the prevailing currents ensured that seals that died close to the shore were likely to be washed up on the beaches of the peninsula. As a result, it was possible to estimate that about 70% of the local population had died, constituting about one-third of the world population.

Hernández *et al.* detected several phycotoxins in the dead seals, and found high concentrations of the dinoflagellate *Alexandrium minutum* in the coastal waters. Other workers<sup>4</sup> have found evidence that the same toxins had entered the nervous system of some of the dead seals. Nevertheless, we cannot be certain that they were the cause of death, because there are no data on lethal or background levels of these toxins in seals.



Figure 1 Bodies on the beach — dead Mediterranean monk seals on the shore of the Cap Blanc peninsula.

Moreover, the algal concentrations recorded are not consistent with a full-blown algal bloom, and some of the toxins that were detected are produced by another species (*Gymnodinium catenatum*), which was only present in small numbers. If there was a bloom, it must have occurred some distance from the seal caves themselves. Algal toxins normally act very quickly, and so it is necessary to explain how and why most of the affected seals returned to the Cap Blanc peninsula to die.

These new results do not discredit the earlier findings. Osterhaus *et al.*<sup>1</sup> isolated virus from some of the dead seals, so there can be little doubt that a morbillivirus was circulating at the time of the die-off. But if it was responsible for all or most of the deaths, it must have acted in a way unlike that observed in other die-offs in which morbillivirus has been implicated, because animals died quickly with few, if any, overt signs of disease<sup>5</sup>. Whether or not it caused the mass mortality, the isolated virus is unlikely to be specific to the monk seal because this species is too scarce and widely dispersed to harbour any normal morbillivirus in an endemic state. The most likely source of the infection was a local dolphin population, because the isolated virus is closely related to one which caused the large-scale deaths of striped dolphins, *Stenella coeruleoalba*, in the Mediterranean in 1990 and 1991 (ref. 6). This is worrisome — it suggests that the threat to endangered species from diseases and parasites carried by more abundant species<sup>7</sup> may come from members of different taxonomic orders, as well as close relatives.

In fact, both agents may have been involved, because this is not the first mass mortality of marine mammals in which an infectious disease and algal toxins have been implicated. In the 1820s, Benjamin Morrell recorded the death of half a million Cape fur seals (*Arctocephalus pusillus*) along the coast of what is now Namibia and attributed it to 'pestilence', although it was subsequently proposed<sup>8</sup> that toxic algae were a more likely cause. The death of about 50% of the bottlenose dolphin (*Tursiops truncatus*) population off the eastern seaboard of the United States in 1987 was first attributed to algal toxins and later to a morbillivirus<sup>9</sup>. Even the mass mortality of harbour seals (*Phoca vitulina*) in the North Sea during 1988 was originally thought to have been caused by a bloom of toxic alga, although there is now ample evidence that a morbillivirus was responsible.

Unfortunately, the Western Sahara population will probably experience more such events, although their time course will depend on which agent is involved. Morbilliviruses are usually highly infectious, so most of the surviving seals have probably been exposed to the virus and will therefore be resistant to future infection for at least the

next seal generation<sup>5</sup>.

By contrast, toxin-producing dinoflagellate blooms may be a relatively new phenomenon in this region, and one to which the seals have had no chance to adapt. *Gymnodinium catenatum* was, until recently, confined to the east coast of the United States, but since 1970 it has become much more widely distributed<sup>10</sup>, possibly because its cysts are easily transported in the ballast water of ships. Previous exposure to algal toxins provides no protection and so another bloom off the coast of the Western Sahara could lead to a new round of seal mortality. This has led to suggestions that some seals from Cap Blanc should be re-introduced into other parts of the species' Atlantic range, such as the Canary Islands, to spread the risk. But this kind of translocation is itself a risky process, the costs and benefits of which cannot yet be evaluated<sup>5</sup>.

The controversy over what killed the monk seals demonstrates the practical and conceptual problems involved in investigating wildlife mortalities<sup>7</sup>. Without baseline

information on the prevalence of virus antibodies, or on levels of algal toxins in seals and their fish prey before the mortality, it is impossible to make a definitive diagnosis. The events at Cap Blanc underline the need for conservation strategies to take account of the vulnerability of endangered species to stochastic events<sup>11</sup> which may be caused by many factors acting in isolation or together. □

John Harwood is in the School of Environmental and Evolutionary Biology, University of St Andrews, St Andrews, Fife KY16 8LB, UK.  
e-mail: jh17@st-andrews.ac.uk

1. Osterhaus, A. *et al.* *Nature* **388**, 838–839 (1997).
2. Hernández, M. *et al.* *Nature* **393**, 28–29 (1998).
3. Gonzalez, L. M., Aguilar, A., Lopez-Jurado, L. F. & Grau, E. *Biol. Conserv.* **80**, 225–233 (1997).
4. Costas, E. & López-Rodas, V. *Vet. Rec.* (in the press).
5. Harwood, J., Lavigne, D. & Reijnders, P. (eds) *IBN Sci. Contrib.* 11 (DLO Inst. Forestry and Nature Res., Wageningen, 1998).
6. Domingo, M. *et al.* *Nature* **348**, 21 (1990).
7. McCallum, H. & Dobson, A. *Trends Ecol. Evol.* **10**, 190–194 (1995).
8. Wyatt, T. J. *Conseil Int. Explor. Mer* **39**, 1–6 (1980).
9. Lipscomb, T. P., Schulman, F. Y., Moffatt, D. & Kennedy, S. J. *Wildlife Dis.* **30**, 567–571 (1994).
10. Anderson, D. M. *Sci. Am.* **271**, 62–68 (1994).
11. Mangel, M. & Tier, C. *Ecology* **75**, 607–614 (1994).

## Bacterial chemotaxis

# United we sense ...

N. Barkai and S. Leibler

Biological systems show remarkably sensitive responses to a wide range of environmental signals. The visual system, for example, can be excited by a single photon, yet can also detect contrasts over a 10,000-fold range of light intensities. How is this impressive combination of sensitivity and wide dynamic range achieved?

On page 85 of this issue<sup>1</sup>, Bray, Levin and Morton-Firth describe how they have studied this question in the context of bacterial chemotaxis, a system that has emerged as a prototype for investigating mechanisms by which cells sense and process information about their surroundings. Bacteria such as *Escherichia coli* can sense chemical attractants over a concentration range of several orders of magnitude, responding even to small signals in which a minute fraction of the chemotaxis receptors bind ligand.

How the bacteria achieve this is a mystery. From their theoretical work, Bray and collaborators offer a solution in which sensitivity and dynamic range are achieved by adaptive receptor clustering. According to this view, the ligand-induced change in the signalling activity of the receptor propagates to a large number of neighbouring receptors, thereby amplifying the effect of a binding event. The authors propose the new idea that, to preserve the dynamic range, the degree of clustering adapts to the external stimuli.

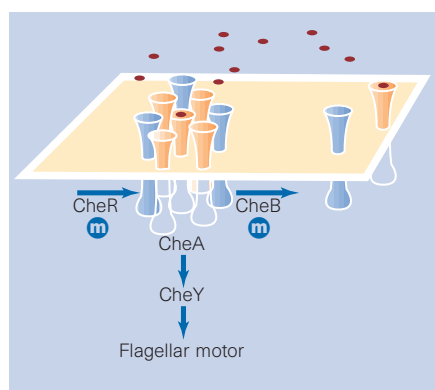
This is an intriguing prospect. Lateral signalling between receptors has been studied

intensively in various other signal-transduction systems. For instance, ligand binding causes receptors for epidermal growth factor to assemble, which allows autophosphorylation of their cytoplasmic domains and thus transmission of the signal downstream. In addition, the different members of this receptor family form heterodimers or higher-order complexes that are thought to be important for diversifying the biological responses specified by the growth factors<sup>2,3</sup>. Another example is the dynamic clustering of T-cell receptors, which seems to play a key role in sustained T-cell signalling<sup>4</sup>.

In *E. coli*, clustering of the chemotactic receptors on the cell membrane was observed several years ago<sup>5</sup>. Hints as to its physiological significance came from genetic experiments<sup>6</sup>, and *in vitro* findings which showed that the biochemical activity of the receptors (kinase activation) involves the formation of large complexes of their signalling domains<sup>7</sup>. What Bray *et al.* have now done is to perform a theoretical analysis, characterizing the interplay between receptor clustering, sensitivity and dynamic range in bacterial chemotaxis.

Rapid migration of bacteria towards favourable chemical environments was first reported over a hundred years ago and has since been a subject of intensive study<sup>8</sup>. In a homogeneous medium, bacteria appear to move in a 'random walk' fashion: long, approximately straight 'runs' are interrupted by short 'tumbles' in which a new





**Figure 1** The chemotactic network. Attractant or repellent molecules bind to the chemotactic receptors that span the bacterial membrane, and induce a conformational change (shown here as a change from blue to orange). This is translated into a change in swimming behaviour through modification of the activity of an intracellular kinase, CheA. CheA modifies a messenger protein, CheY, which binds to the flagellar motor and induces tumbling. Adaptation involves reversible methylation (m) of the receptors by opposing enzymes CheR and CheB (an enzyme activated by CheA). Bray *et al.*<sup>1</sup> propose that the ligand-induced activity change of the receptor propagates in a cluster of neighbouring receptors, a process that amplifies the signal and may account for the high gain of the chemotactic response. The authors suggest that, to preserve the dynamic range, both clusters and single receptors coexist on the cell membrane.

direction of motion is chosen. To achieve chemotaxis, a bacterium that is moving in a favourable direction along a chemical gradient suppresses tumbling events, and thus migrates, on average, towards attractant and away from repellent. Although tumbling frequency is modulated by changes in chemical concentrations, its steady-state value is independent of the concentration itself. This property is called exact adaptation. Among other effects, it is believed that adaptation allows bacteria to respond to gradients over a wide range of background chemical levels.

The biochemical network that mediates the chemotactic response is relatively simple<sup>8</sup>. Sensory receptors modulate the activity of an associated kinase that phosphorylates a messenger protein, which, by interacting with flagellar motors, induces tumbling (Fig. 1). Binding of an attractant, such as aspartate, to the receptor suppresses tumbling by inhibiting the kinase activity. Adaptation is brought about by a feedback mechanism involving reversible methylation of the receptors.

The efficiency of the chemotactic response is evident in its high sensitivity — binding of attractant to less than 1% of the receptors is sufficient to induce measurable changes in the tumbling frequency<sup>1,9</sup>. This sensitivity is also remarkable in view of the low abundance of some receptors that func-

tion through the same phosphorylation cascade. For instance, addition of ribose can suppress tumbling completely, even though only a small fraction of all receptors is sensitive to ribose<sup>1</sup>. The origin of this high sensitivity, or gain, remains unknown and it is perhaps one of the most striking aspects of the chemotactic response. In fact, none of the quantitative models of the chemotactic network<sup>10–13</sup> has accounted for the observed gain, although they have successfully explained several aspects of the response, including adaptation. This is particularly distressing in view of experiments that quantitatively measured the interaction between the messenger proteins and the flagellar motor and found a relatively low degree of cooperativity (ref. 14, and U. Alon *et al.*, personal communication).

What, then, is the origin of the reported high sensitivity of the chemotactic response. There are several ways in which the observed gain might be generated by signal processing within the biochemical network. It could simply result from an as yet unidentified, highly cooperative protein–protein interaction, or from network amplification processes such as those observed in other phosphorylation cascades. Such mechanisms operate downstream of the receptor–ligand binding and amplify the signalling activity of the receptors. High sensitivity then requires the steady-state activity of the receptor to be adjusted with the ‘edge of sensitivity’ of the amplifying process.

Bray and collaborators<sup>1</sup> consider a different type of mechanism, where the high gain is the result of interactions between receptors. Here, the ligand-induced change in conformation, which is responsible for the transmission of the tumble signal to the motor, can spread between receptors. Such ‘infection’ amplifies the effect of ligand binding, and may account for the observed gain. In principle, this mechanism does not restrict the level of receptor signalling activity, as this activity is now the output of the amplification process. Bray *et al.* demonstrate, however, that whereas a high degree of receptor ‘infection’ increases the gain of the system, it also severely reduces the dynamic range, because the receptors will already be saturated at low chemical concentrations. Such an apparent incompatibility between high sensitivity and wide dynamic range is a general problem for many sensory systems; the solution is indeed provided through adaptation.

Bray *et al.* propose the existence of an additional form of adaptation in bacterial chemotaxis: that the degree of receptor clustering depends on the concentration of ligand. Although the detailed molecular mechanism that accounts for such adaptation needs to be elucidated, the proposed dependence fits in well with observations on receptor function in other systems<sup>2–4</sup>. If confirmed by biochemi-



#### 100 YEARS AGO

The theory of the origin of sleep which has gained the widest credence is the one that attributes it to anaemia of the brain. ... It has been supposed, but without sufficient evidence to justify the supposition, that this anaemia of the brain is the cause and not the sequence of sleep. The idea behind this supposition has been that, as the day draws to an end, the circulatory mechanism becomes fatigued, the vasomotor centre exhausted, the tone of the blood vessels deficient, and the energy of the heart diminished, and thus is the circulation to the cerebral arteries lessened. ... The alternative theories, which have been suggested to account for the onset of sleep, may be classed as chemical and histological. ... It is held possible that the dendrites or branching processes of nerve cells are contractile, and that they, by pulling themselves apart, break the association pathways which are formed by the interlacing or synapses of the dendrites in the brain. Ramón y Cajal, on the other hand, believes that the neuroglia cells are contractile, and may expand so as to interpose their branches as insulating material between the synapses formed by the dendrites of nerve cells.

From *Nature* 5 May 1898.

#### 50 YEARS AGO

As I look at a living organism, I see reminders of many questions that need to be answered. Not all these questions are obviously important, nor would their answers be useful – but we want them answered. Thomas Wright in 1601 said, “Nothing is so curious and thirsty after knowledge of dark and obscure matters as the nature of man” – of scientific men especially, he might have said. What is skin, fingernail? How do fingernails grow? How do I feel things? How are nerves built and how do they function? How do I see things? ... The basic answers to all these questions are not to be found in books. Even though Chaucer said

For out of olde felde, as men seith,  
Cometh al this newe corn fro yeer to yere;  
And out of old bokes, in good feith,  
Cometh al this newe science that men lere,  
he was before long corrected by Francis Bacon: “Books must follow sciences, and not sciences books”. – Linus Pauling  
From *Nature* 8 May 1948.



cal and structural data, this adaptation would account for both high sensitivity and wide dynamic range — and this, in principle, without need of 'fine-tuning' mechanisms<sup>10</sup>.

Research into bacterial chemotaxis is thus entering a new stage. A detailed description of molecular components and their interactions is now available, and bacterial behaviour can be assessed quantitatively. The challenge is to connect these two levels of description through an intermediate-level, quantitative 'system analysis'. By addressing the issue of the system's sensitivity and its dynamic range, Bray *et al.* have taken a step in this direction. Indeed, bacterial chemotaxis may become a prototype in the search for the underlying design principles of biochemical networks. □

N. Barkai and S. Leibler are in the Departments of Physics and Molecular Biology, Princeton

University, Princeton, New Jersey 08544, USA.

e-mail: barkai@phoenix.princeton.edu

1. Bray, D., Levin, M. D. & Morton-Firth, C. J. *Nature* **393**, 85–88 (1998).
2. Pinkas-Kramarski, R. *et al.* *EMBO J.* **16**, 2452–2467 (1996).
3. Riese, D. J. & Stern, D. J. *BioEssays* **20**, 41–48 (1998).
4. Germain, R. N. *Curr. Biol.* **7**, R640–R644 (1997).
5. Maddock, J. R. & Shapiro, L. *Science* **259**, 1717–1723 (1993).
6. Stock, J. B. *Science* **274**, 370–371 (1996).
7. Liu, Y., Levit, M., Lurz, R., Surette, M. G. & Stock, J. B. *EMBO J.* **16**, 7231–7240 (1997).
8. Stock, J. & Surette, M. in *E. Coli and S. Typhimurium: Cellular and Molecular Biology* (ed. Neidhart, F. C.) 1103–1129 (Am. Soc. Microbiol., Washington DC, 1996).
9. Block, S. M., Segall, J. E. & Berg, H. C. *Proc. Natl Acad. Sci. USA* **83**, 8987–8991 (1986).
10. Bray, D., Bourret, R. B. & Simon, M. I. *Mol. Biol. Cell* **4**, 469–482 (1993).
11. Hauri, D. C. & Ross, J. A. *Biophys. J.* **68**, 708–722 (1995).
12. Spiro, P. A., Parkinson, J. S. & Othmar, H. G. *Proc. Natl Acad. Sci. USA* **94**, 7263–7268 (1997).
13. Barkai, N. & Leibler, S. *Nature* **387**, 913–917 (1997).
14. Scharf, B. E., Fahrner, K. A., Turner, L. & Berg, H. C. *Proc. Natl Acad. Sci. USA* **95**, 201–206 (1997).

## Zeolites

# Organic groups cling to the pores

Edward J. Creighton

The need for sophisticated catalysts is driven by the need to reduce waste emission from chemical synthesis. In manufacturing complicated chemicals, especially pharmaceuticals, the amount of waste often exceeds the amount of product. Therefore, more selective and cleaner conversions are increasingly desired<sup>1</sup>. On page 52 of this issue<sup>2</sup>, Jones *et al.* present a new class of shape-selective catalyst, by synthesizing a zeolite functionalized with organic groups. Zeolites made in this way might be used to replace aluminium chloride and other wasteful catalysts.

Zeolites are aluminosilicates with well-defined pores, and often cavities, of molecular dimensions. Their shape selectivity comes in three forms<sup>3</sup>. In the first of these,

the pores can act as a molecular sieve, preventing large molecules from entering — this is called reactant selectivity. The paper by Jones *et al.* is an excellent example<sup>2</sup>: in the acetalization of carbonyl compounds with ethylene glycol (catalysed by phenyl-sulphonic-acid groups tethered in the micropores of zeolite beta), cyclohexanone is readily converted, whereas 1-pyrenecarboxaldehyde, which is too large to enter the pores of the modified zeolite, is not (Fig. 1). A subtler form of reactant selectivity is based on diffusion: molecules with higher diffusivities than competing reactants will generally react faster. The second type of selectivity, 'product selectivity', occurs when only those products with appropriate dimensions can escape from the interior of the zeolite.

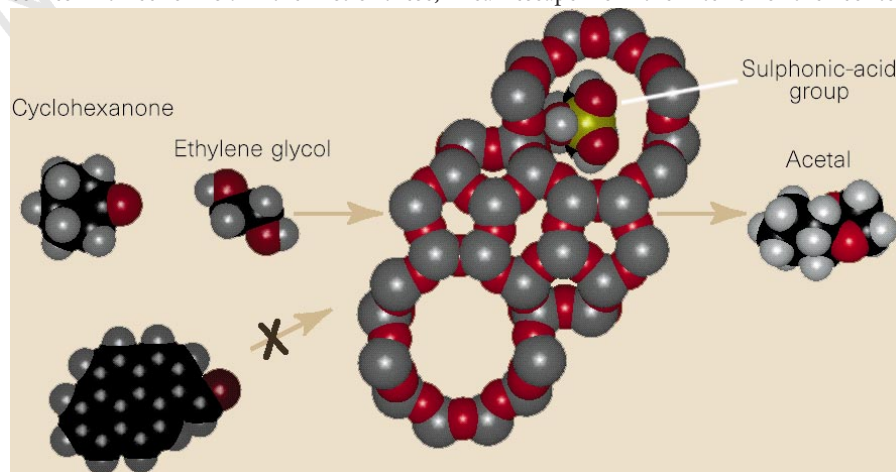


Figure 1 Reactant selectivity: a zeolite acting as a molecular sieve. Cyclohexanone and ethylene glycol can easily reach the sulphonic-acid group in the micropores of zeolite beta and so can react to form acetal. The larger molecule, 1-pyrenecarboxaldehyde, cannot enter the pores and will remain unaffected. Jones *et al.*<sup>2</sup> have developed a new way to make zeolites with functionalities such as this sulphonic-acid group.

Finally, 'restricted transition-state selectivity' involves the prevention or retardation of certain reactions because their corresponding transition states require more space than is available inside the zeolite.

Owing to these remarkable properties, zeolites are used on a large scale as solid-acid catalysts in the petroleum and petrochemical industries — in hydrocracking, fluidized catalytic cracking, catalytic dewaxing, aromatics processing, reforming, and making synthetic fuels, for example<sup>4,5</sup>. In many cases they are chosen because of their high thermal stability, high acidity and unique shape-selective properties.

Active sites can also be deliberately introduced into zeolites. This can be accomplished by substituting transition metals for aluminium in the zeolite framework, by introducing transition-metal complexes in the zeolite cavities as 'ship-in-a-bottle' systems, by simple ion-exchange, or by grafting and tethering of organometallic complexes. In this way, shape-selective catalysts can be created for oxidation, hydrogenation, acid or base catalysis, and a range of other functions<sup>5,6</sup>.

Jones *et al.*<sup>2</sup> describe a new class of zeolite-based catalysts. They have managed to synthesize zeolite beta functionalized with organic groups, simply by adding the organosilicon compound phenylethyl-trimethoxysilane to their synthesis gel (which also contains a silicon source, water, an organic structure-directing agent (SDA) and fluoride ions as a mineralizing agent). After the zeolite forms, the organic phenylethyl groups serve as anchors for the sulphonic-acid groups — the actual acid sites — which are introduced after removal of the pore-filling organic SDA. Another way of introducing organic anchors would be grafting with organosilanes, but this would probably result in a less even distribution of organic functionalities throughout the zeolite.

The method can only be applied to zeolites that can be synthesized without an organic SDA, or when the SDA can be dissolved out. This reduces the general applicability of the method, as few zeolites can be synthesized without an SDA, and although the use of solution-extractable (and thus recyclable) SDAs is increasing, most commercially produced zeolites still need to be calcined (heated) to make the micropores accessible. Such heating would destroy the phenylethyl groups of Jones *et al.*

Another point of concern is the size of the organic functionality. Benzene, for instance, has a kinetic diameter of 0.58 nm, which means that it will already fill the pores of zeolite beta (which measure 0.64–0.76 nm). So tethered phenyl-sulphonic-acid groups larger than benzene can block the pores entirely. To avoid this, the more spacious zeolites are to be preferred, with

pore diameters of about 0.7 nm (with a circumference of 12 oxygen atoms) and with either a three-dimensional intersecting channel system or large cages. The only two known zeolites that fulfil these requirements are zeolite beta and zeolite Y.

The latter is more interesting as a catalyst because it can be synthesized easily without an organic SDA and it has a three-dimensional channel system, defined by interconnecting cages with a diameter of about 1.3 nm, which can serve as microreactors. At present, 90,000 tonnes a year of zeolite Y is used in the petroleum industry for acid catalysis (cracking). The industrial applications of zeolite beta are rapidly increasing, but it cannot be made on a commercial scale without an organic SDA, which is expensive and can't yet be recycled.

Nevertheless, Jones *et al.*<sup>2</sup> demonstrate the potential of organic functionalized zeolites in shape-selective catalysis. It would be interesting to test the catalyst in Friedel-Crafts alkylation/acylation. In this field, solid-acid catalysts are needed to replace aluminium chloride, which is highly corrosive and produces large amounts of waste. Aluminium-containing acid zeolites, which are hydrophilic, can be used, but an imbalance between the reactants in the micro-

pores, resulting from differences in polarity, often limits the conversion. This problem might be solved by using the more hydrophobic organic functionalized all-silica lattice of Jones and colleagues.

This new method could also be extended to other types of catalysis, by finding appropriate compounds to add to the synthesis gel. The next step might be introducing even stronger acid or base sites to a zeolite, or anchoring metal-ligand complexes that catalyse homogeneous reactions, such as selective oxidation or hydrogenation (although stabilizing these catalysts against leaching of the metal could be a problem). A more distant goal is the synthesis of solid chiral catalysts for asymmetric reactions, such as enantioselective epoxidation, hydrogenation and Lewis-acid-catalysed rearrangements. □

Edward J. Creighton is at the Shell Research and Technology Center, Amsterdam, PO Box 38000, 1030 BN Amsterdam, The Netherlands.

e-mail: eddy.j.creyghton@opc.shell.com

1. Downing, R. S., van Bekkum, H. & Sheldon, R. A. *Cattech* **2**, 95–109 (1997).
2. Jones, C. W., Tsuji, K. & Davis, M. E. *Nature* **393**, 52–54 (1998).
3. Csicsery, S. M. *ACS Manag.* **171**, 680–713 (1976).
4. Maxwell, I. E. & Stork, W. H. J. *Stud. Surf. Sci. Catal.* **58**, 571–630 (1991).
5. Venuto, P. B. *Microporous Mater.* **2**, 297–411 (1994).
6. Venuto, P. B. *Stud. Surf. Sci. Catal.* **105**, 811–851 (1997).

## Hybrid genetics

# Transposons unbound

Margaret G. Kidwell and Damon R. Lisch

Large-scale genomic sequencing has revealed that a huge fraction of many eukaryotic genomes consists of transposable elements (transposons). These are mobile DNA sequences — ranging in size from several hundred to many thousands of base pairs — that can 'jump' from one chromosomal location to another. At least 35% of the human genome is made up of transposon DNA, and unfettered transposon activity can lead to rapid genomic disruption resulting from mutations and chromosome rearrangements.

Not surprisingly, a number of mechanisms have evolved to regulate these potentially disruptive invaders. One such mechanism is DNA methylation — a modification of cytosine residues that leads to transcriptional silencing of targeted DNA sequences. Indeed, the main role of methylation in primates was recently claimed<sup>1,2</sup> to be defence against the activity of parasitic transposable elements. This is in sharp contrast to the long-held view that the function of methylation is to regulate normal development.

A dramatic illustration of the effect of lifting the methylation-induced protection against transposable elements is provided by Waugh O'Neill and colleagues<sup>3</sup> on page 68 of this issue. An interspecific hybrid between

two species of Australian wallaby (*Macropus eugenii* and *Wallabia bicolor*) shows genome-wide undermethylation compared with its parental species. The authors found that this hybrid also had atypically extended centromeres (the region of the chromosome to which the spindle fibres attach), and that the additional centromeric material consisted

of an unmethylated, endogenous retroviral element (retroelement), amplified many times (Fig. 1). Extended centromeres were not found in either parental species. Moreover, preliminary analysis of hybrids between other species of wallaby suggested that this process might be a general feature of hybridization between these species.

The insertion profile of the invading elements is consistent with previous observations from species as divergent as the fruit fly *Drosophila melanogaster* and maize. The regions around centromeres contain many transposable-element sequences, suggesting that these regions are a 'safe haven' for transposons that have escaped inactivation or elimination<sup>4,5</sup>. The work of Waugh O'Neill *et al.* now provides evidence that such a distribution of elements can result from specific targeting of particular chromosomal regions, rather than long-term selection on a population of randomly distributed elements.

The massive amplification of retroelements in a single generation provides an impressive example of the capacity of transposable elements to change the structural architecture of host genomes without disrupting the normal functioning of somatic cells. The results provide a fascinating glimpse into how a parasitic element can invade a genome. But is there any evidence that retroelements are involved in the evolution of species karyotypes (the chromosomal make-up of a species), or the evolution of reproductive isolation and speciation? For that to be the case, changes mediated by the transposable element would have to be heritable. In this regard, the degree of activity of transposable elements in germ cells — the cells that are passed on to future generations — is relevant. The degree to which DNA methylation normally limits colonization of germline genomes by retroelements is hotly debated<sup>1,6</sup>, so the relevance of the present

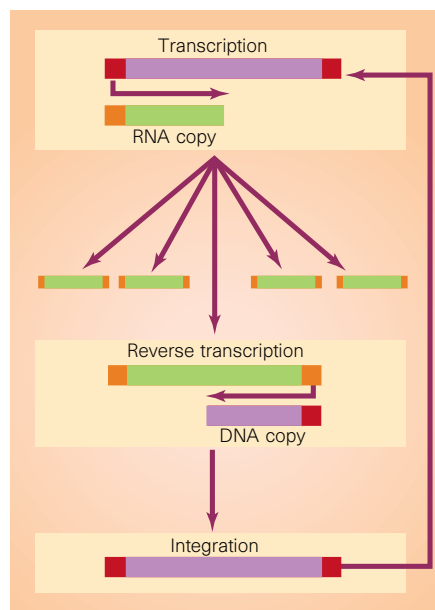


Figure 1 Replicative transposition of retroviral elements (retroelements). Such transposition can result in rapid amplification of the retroelement, as observed by Waugh O'Neill *et al.*<sup>3</sup> in hybrids between two species of Australian wallaby. RNA copies of an integrated element are produced through transcription, then some are reverse-transcribed back to DNA. The DNA copies are re-integrated into the host's chromosomes, and the cycle can begin again. This process can amplify the copy number of the retroelement, potentially leading to a massive increase in copy number over a short period of time. The boxed regions in red and orange represent direct repeats, which are necessary for retroviral integration.

results for germ cells is not known. Furthermore, because the hybrids were sterile, Waugh O'Neill *et al.* could not address this question, either in the individual concerned or in its offspring. Fertile hybrids and their progeny will need to be examined to study the evolutionary significance of these results.

Although it is rare in animals, the evolution of new species from interspecific hybrids is well documented in plants. Often it is associated with polyploidy (a greater than diploid number of chromosomes) and asexual reproduction. According to population-genetic theory, speciation is possible after recombination occurs in hybrids. In this mode of speciation, hybridization is followed by inbreeding and breakdown of previously linked segments<sup>7</sup>. If selection for viable and fertile recombinant phenotypes is successful, and these are reproductively isolated from the parents, the recombinants may be stabilized and a new species may evolve. Rare, chromosomal rearrangements that lead to reduced fitness and might mediate reproductive isolation are unlikely to survive in large populations. In very small populations, however, the chance for the rearrangement to be fixed by genetic drift is greatly improved by increased mutation rates<sup>8</sup>.

Mobilization of transposons is associated with high rates of karyotypic change in

*Drosophila*<sup>9</sup> and yeast<sup>10</sup>, and there is little doubt that karyotypic changes can be associated with speciation. Moreover, transposon activity is often associated with increased mutation rates. Therefore, although the relationship between phenomenology and evolutionary significance remains ill defined, transposons seem likely to be involved in the evolution of new species. Waugh O'Neill *et al.*<sup>3</sup> provide a glimpse of the enormous potential of transposable elements for mediating rapid and large-scale changes in genome composition and architecture. What is less evident is the degree to which transposons have contributed to the evolution of their hosts — their fingerprints are everywhere, but the culprits remain elusive. □

Margaret G. Kidwell and Damon R. Lisch are in the Department of Ecology and Evolutionary Biology, and the Department of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA.  
e-mail: kidwell@azstarnet.com.

1. Yoder, J., Walsh, C. & Bestor, T. *Trends Genet.* **13**, 335–340 (1997).
2. Henikoff, S. & Matzke, M. A. *Trends Genet.* **13**, 293–295 (1997).
3. Waugh O'Neill, R. J., O'Neill, M. J. & Marshall Graves, J. A. *Nature* **393**, 68–72 (1998).
4. Jiang, J. *et al. Proc. Natl Acad. Sci. USA* **93**, 14210–14213 (1996).
5. Sun, X., Wahlstrom, J. & Karpen, G. *Cell* **91**, 1007–1019 (1997).
6. Bird, A. *Trends Genet.* **13**, 469–472 (1997).
7. Templeton, A. R. *Annu. Rev. Ecol. Syst.* **12**, 23–48 (1981).
8. Fontdevila, A. *Genetica* **86**, 247–258 (1992).
9. Lim, J. K. & Simmons, M. J. *BioEssays* **16**, 269–275 (1994).
10. Codon, A. C., Benítez, T. & Korhola, M. *Curr. Genet.* **32**, 247–259 (1997).

## Ecology

# The forest fragment classic

Stuart L. Pimm

Twenty years ago, Thomas Lovejoy from the Smithsonian Institution in Washington DC engineered a unique ecological experiment from a chance alignment of human demography, science and politics. A rapid influx of people into one of the planet's most sparsely populated areas — the Amazon — was resulting in the clearance of rainforest at unprecedented rates. Complete deforestation would eliminate forest-dependent species. But, in practice, isolated forest fragments of various sizes would probably remain, and the uncut forest would have convoluted edges.

The laws of the Brazilian state of Amazonas required that forest clearing leave 50% of the forest intact. (Those laws now require it to be 80%.) Lovejoy coordinated an international project between the Instituto Nacional de Pesquisas da Amazônia (INPA) in Manaus and US-based partners to ensure that forest clearing in an area north of Manaus would leave neat squares of 1, 10 and 100 hectares of forest (Fig. 1, overleaf). These fragments would enjoy long-term protection by presidential decree, and, at a recent meeting\*, ecologists assessed the results

from two decades of monitoring them.

Two sets of questions had originally generated controversy. First, would the observations of island biogeography (whereby small, oceanic islands have fewer species than larger ones, and islands distant from mainland sources of species have fewer species than closer ones) extend to forest 'islands' surrounded by a 'sea' of cattle pastures? If so, how large, and how near to source areas, would forest reserves need to be to retain most of their species? Second, although forest edges are palpably hotter, drier and visibly have more tangled vegetation than forest interiors, how far into the forest would these and other 'edge effects' penetrate?

It is no longer controversial to say that small, isolated forest fragments lose species — the rapid pace of temperate and tropical deforestation has supplied many examples worldwide<sup>2</sup>. At Manaus, various international teams joined with INPA scientists to show that the diversity of palms, euglossine bees, shade-requiring butterflies, dung beetles, termites, birds and primates all declined in the fragments. Some species were lost from the smaller fragments within a few years of their isolation from the once-continuous forest<sup>3</sup>.

The mechanisms behind these losses were sometimes those posited by the theory of island biogeography. Primates, for example, have large home ranges, and fragments were too small to house viable populations. Army ants also range widely, bivouacking each day rather than foraging from a fixed nest. Antbirds follow army ants, scavenging the insects they disturb, and, not surprisingly, antbirds are now rare even in the 10-hectare plots (P. Souffer and J. Stratford, Southeastern Louisiana Univ.). Other mechanisms involve the poorer establishment of forest species in smaller fragments. For instance, transplanted seedlings of *Copaifera multijuga* survive better in larger patches or in continuous forest (M. Elias and I. Ferraz, INPA). Interactions between species are also affected. In the small fragments, for example, the abundance of euglossine bees has declined tenfold — have plants suffered from the loss of these once-common pollinators?

The island biogeographic model fails for some groups, however. Unlike true islands — where ocean surrounds land — forest species can use the land surrounding forest islands to different degrees, and some species flourish in disturbed habitats. The number of leaf-cutter ants, for example, increased 30-fold in cleared areas, spilling over into the forest fragments (H. Vasconcelos, INPA). Moreover, the number of species of frog in the fragments also increased. This was unexpected, but many species that were thought to depend on rainforest seemed to thrive in farm ponds (M. Toucher, Univ. Canterbury). To manage the biological diversity of the planet's ever-more fragmented forests, we therefore need to understand why fragmentation affects some species but not others.

Previous five-year meetings to assess the project's progress were critical of its failure to look at the changes in plant communities. The taxonomic problems are formidable — for example, the fragments have more families of tree than Britain has species. But William and Susan Laurance, Leandro Ferreira and Patricia Delamonica (INPA) have overcome these difficulties, setting up a herbarium with 45,000 identified specimens. These data, combined with the results from control plots within the unbroken forest, allowed the project to address another question that was not anticipated at its outset: is the Amazon rainforest a source or a sink of global carbon? Gas-exchange studies above the forest canopy indicate that the Amazon basin could be a large carbon sink<sup>4</sup>. But results from the Manaus project suggest that there is considerable regional variation in carbon accumulation (O. Phillips, Univ. Leeds; W. Laurance). There is considerable temporal variation, too. The current El Niño has deprived the Amazon of its normal rainfall, and, in the control plots, tree mortality was twice the normal levels. So, this season the Amazon will fix less carbon than in

\*Biological Dynamics of Forest Fragments Project, Five-Year Review, Manaus, Brazil, 15–18 March 1998.



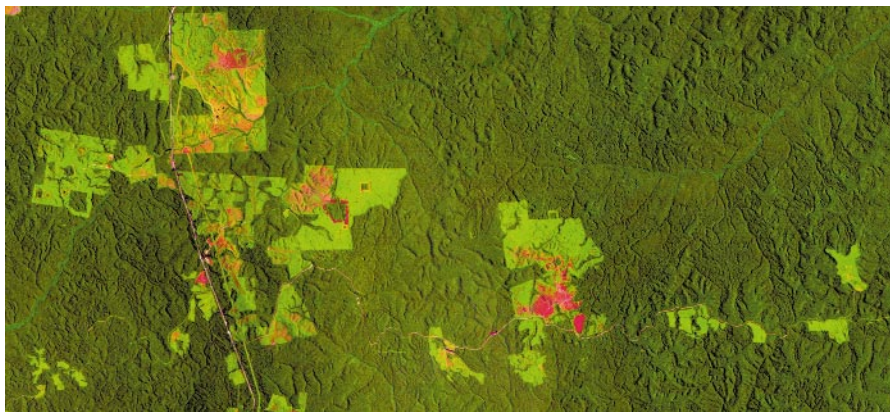


Figure 1 Forest clearings in the Amazon, in an area north of Manaus, Brazil. Clearings show as pale green among the uncut, darker green forest. Within the clearings, square forest islands of different sizes remain. These, plus control plots within the continuous forest, constitute the forest-fragment experiment. The pink areas are bare soil or recently cleared areas, including those that must be repeatedly cleared around the experimental plots. The clearing at the far right of the image is about 3 km east to west and 5 km north to south.

average years (W. Laurance; B. Williamson, Louisiana State Univ.).

The second controversial question raised by the project concerned the forest edges. The project could not have anticipated the recent history of forest clearing, and the relative amounts of forest islands and forest edges that it produced. Depending on definition, the Amazon basin has 3–4 million km<sup>2</sup> of lowland rainforest<sup>5</sup>. Brazil's Instituto Nacional de Pesquisas Espaciais (INPE) estimated Amazon deforestation at 21,000 km<sup>2</sup> per year<sup>6</sup> for the 1980s, and, since then, the annual estimates range from 50 to 150% of that amount<sup>6</sup>. INPE, however, does not include the amounts of forest left behind as islands, or the amounts near forest edges.

A study of satellite images by David Skole (Univ. New Hampshire) and Compton Tucker (NASA) broadly agrees with these rates of deforestation<sup>6</sup>. Deforestation has left little forest behind, and the laws on leaving a specified fraction uncut have not been widely respected. Skole and Tucker found that little of what does remain is in isolated forest islands — only about 1,000 km<sup>2</sup> per year. By comparison, the amount of forest edge is huge. Large roads bisect the forest, small roads feed off them perpendicularly at regular intervals, and yet-smaller tracks feed off these. Skole and Tucker asked how much forest was within 1 km of these numerous edges, and the answer was an area one and a half times larger than that of the cleared forest.

Results from the Manaus project show large changes in species composition, population dynamics and carbon flux along this huge area of new forest edge. There are sharp increases in tree mortality and damage because of greater exposure to wind, penetrating up to 300 m (perhaps further) from the forest edges. Fragments of less than 36 hectares are all 'edge', and, within them, about 10% of the plant biomass was lost within two to four years of isolation<sup>7</sup>. Lau-

rance and colleagues reported that these changes accompany accelerated tree dynamics — species now grow faster and die younger than before. And although the abundance of lianas increases along the edges, this is not enough to compensate for the loss of tree biomass. As a result, up to 15 million tons of carbon are lost from the forest to the atmosphere from the newly created forest edges in the Amazon. Worldwide, the amount is perhaps ten times that.

With 20 years of hindsight, Lovejoy's second question, about edges, is more important than the first one about forest islands. Tropical deforestation has been so complete that forest islands constitute only a small area. It is understanding the extent to which edge effects penetrate uncut forest, and how these effects are modified by the nature of what surrounds them (secondary forests or cattle pastures), that constitute the pressing ecological questions for Claude Gascon, the project's lead scientist. Of course, at the current rates of deforestation, not even forest edges will remain a century from now. Only with a determined effort might some forest islands remain. If so, the results from this project would provide the best hope for informing their management. □

Stuart L. Pimm is in the Department of Ecology and Evolutionary Biology, Dabney Hall, The University of Tennessee, Knoxville, Tennessee 37996, USA. e-mail: StuartPimm@aol.com

1. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
2. Laurance, W. F. & Bierregaard, R. O. (eds) *Tropical Forest Remnants: Ecology, Management, and Conservation of Fragmented Communities* (Univ. Chicago Press, 1997).
3. Lovejoy, T. E. et al. in *Conservation Biology: The Science of Scarcity and Diversity* (ed. Soulé, M. E.) 257–285 (Sinauer, Sunderland, MA, 1986).
4. Grace, J. et al. *Science* **270**, 778–780 (1995).
5. Skole, D. & Tucker, C. *Science* **260**, 1905–1910 (1993).
6. INPE *Deforestation Rates for the Brazilian Amazon, 1995–1997* (Diário Uniao Oficial de Brasil, Brasília, 1998).
7. Laurance, W. F. et al. *Science* **278**, 1117–1118 (1997).

## Daedalus

### Gravity waving to us

Physicists are still exasperated by the elusiveness of gravitational waves. Laser interferometers and resonating metal bars have so far failed to detect them. Einstein showed that a gravitational field can be detected by its bending of light. A distant galaxy in the same line of sight as a nearer one can form an 'Einstein ring' image around it; the light is deviated in passing through the nearer galaxy's gravitational field. If an Einstein ring is ever found around an empty patch of sky, a clump of dark matter will have been discovered. Similarly, says Daedalus, the light from a distant star or galaxy should be deviated in phase with any gravitational waves through which that light passes. An observer would see the object apparently vibrating back-and-forth in the sky, in time with the gravitational waves.

The effect would be extremely small, and is probably below the resolution of the best modern telescopes. But Daedalus is undismayed. He points out that phase-sensitive detection can extract a periodic signal from vastly greater amplitudes of noise, provided the frequency and phase of the signal are known. Several types of astronomical object, such as spinning pulsars and tightly orbiting binary neutron stars, should radiate gravitational waves at their own rotational frequency. So he wants to train a big telescope on such an object, and study the image of a distant star or galaxy close to it in the sky. The image should be vibrating ever so slightly in time with the rotation frequency of the gravitational source. Demodulate the position-signal of that star with a phase-sensitive detector locked to the rotating source, and the amplitude and direction of the vibration should slowly emerge out of the optical noise.

This elegant scheme only works with gravitational waves of known period. But once the method has been well worked out, it should be possible to select the image of a distant galaxy, and to try decoding its position-fluctuation with a whole range of likely frequencies and phases. The occasional positive result would indicate the presence nearby of some gravitational source with that frequency and phase. The real prize would be to detect several such objects, all vibrating in the same phase as some invisible central source of gravity waves. This would reveal dark matter in the form of two extinct neutron stars, dead and dark but still locked in each other's gravitational embrace, and still waltzing endlessly round each other.

David Jones

# Noticing *Nature*

Over almost 130 years, this journal's appearance has evolved along with changes in design fashions and in the way scientists present their results. The first in a series exploring how science uses visual images.

**Martin Kemp**

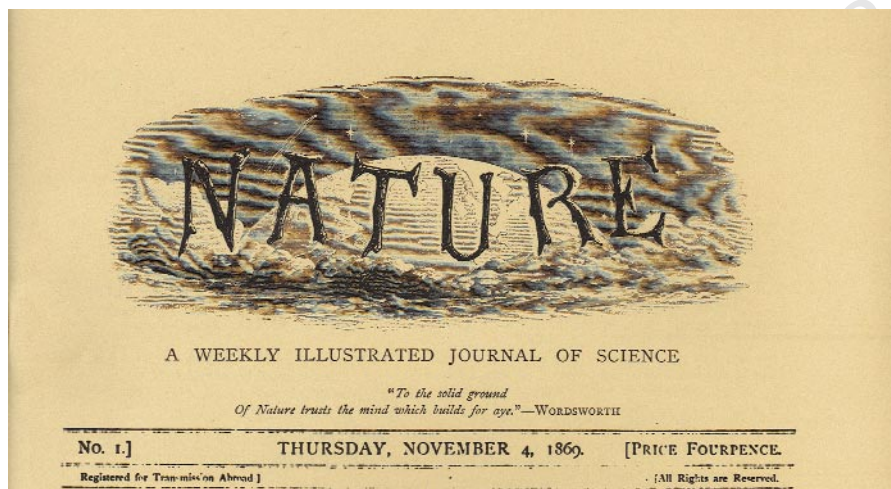
The first issue of *Nature* on 4 November 1869 announced itself as "A Weekly Illustrated Journal of Science". A letter touting for subscriptions in 1998 extols "*Nature's* beautiful presentation" and state-of-the-art visuals: "the 'new-look' *Nature* now fully exploits modern graphics and typography". The journal's emphasis upon the centrality of illustration has remained, but little else is the same. The few constants — most obviously the use of line diagrams — are greatly outweighed by conspicuous changes.

The visual evolution of *Nature* serves to show not only how notions of 'illustration' have been transformed over the 129 years, but also how visual styles for the conduct and broadcasting of the sciences have undergone fundamental change. It is easy, at a distance of historical remove, to discern the visual rhetoric appropriate to 1869. We are less alert to our own period style.

The masthead of the newly emergent *Nature* appeared atop two columns of advertisements, including such delights as "Kemble's Shakespeare Readings", as well as the expected books of science. It displayed planet Earth rising — presumably not sinking — through celestial clouds against a firmament of twinkling stars, across which sped a comet. "NATURE" is spelt out in the twiggly calligraphy of a rustic gazebo and, nestling in the crook of the 'U', Britain is accorded a comfortably secure position in the cosmos of scientific nations.

**As a visual product, it was thoroughly at home on the solid shelves in the library of one of the great Gothic 'town halls' built around 1850 for the conduct of science by the universities.**

This present issue's computerized, full-colour cover sets white or pale-coloured type



Original 1869 masthead: rustic calligraphy, and an image appropriate to a nineteenth-century library.

against the out-of-focus background of a pretty photograph of two hermit crabs in an 'agonistic encounter'. The lettering is seemingly pushed forward onto an ambiguously positioned 'window' in front of the photograph, giving a sense of transparent depth to the cover's flat surface. The image is brisk, up-to-date and snappy, exuding a sense of high-tech mastery without being too blatant and obvious, which would lose the sense of wonder that recent covers strive to evoke. Vivid contemporary design is intended to make the journal look the part, and ensures that it can compete for notice in an age when every publication shouts for visual attention.

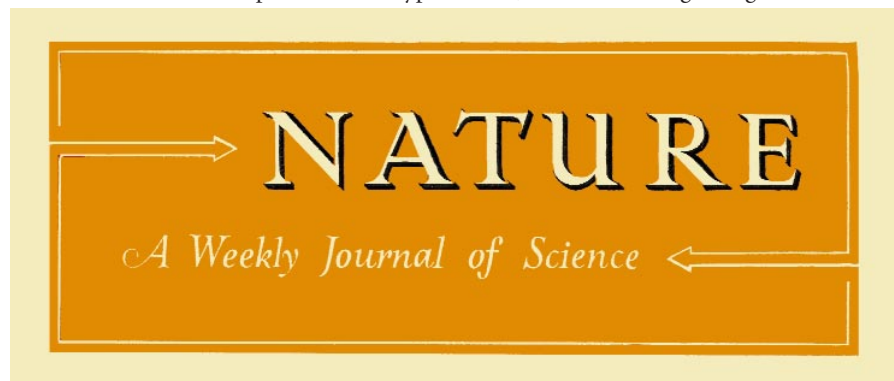
The present design is not the first 'new-look' *Nature*. A 'tidied-up' version of the globe masthead — with a cleaner Roman typeface and more clearly delineated globe which located Britain in an even more central position — surprisingly survived into the late 1950s. As always, there must have been a tension between updating the image and losing 'product identification'. For a long time, the familiar image hung on.

**By contrast, *Nature* prepared to enter the 1960s with a radical redesign, led by a punchier two-colour masthead, hugely simplified in keeping with a more modern image.**

The only non-verbal elements were two purposeful arrows, denoting progress and pointing by implication into the optimistic future. The Wordsworth quotation in the original design survived this revolution, repositioned on the title page, only to disappear quietly in 1963. In 1968, the arrows went too. "NATURE, INTERNATIONAL JOURNAL OF SCIENCE" was printed in the starkest possible typeface and forced itself loudly against the margins of its coloured band.

By 1974, the present, more friendly Garamond typeface in democratically trendy lower case was in place, printed in white, generally against orange, and the advertisements had been replaced by a striking photographic image. The present, glamorously integrated covers of words and images, taking advantage of computer design programs, were initiated in 1993.

The visual styles over the years are very much of a piece with the tone of the texts. The capital 'N' in the Wordsworthian homily — "To the solid ground of Nature trusts the mind which builds for aye" — is much in keeping with the opening essay commissioned from Thomas Huxley, which was largely given over to his translation of "Aphorisms by Goethe". Written around 1780, Goethe's paradoxes of ecstatic wonder are full of what he later admitted was "a sort of Pantheism, to the conception of an unfathomable, unconditional, humorously self-contradictory Being, underlying the phe-



Modern masthead, 1958: punchy, simple, two-colour design, pointing the way to scientific progress.



# NATURE

## INTERNATIONAL JOURNAL OF SCIENCE

Ultra-modern masthead, 1968: stark, uncompromising and fashioned in the 'white heat' of the technological age.

nomena of Nature". Although the first published article, "On the Fertilisation of Winter-Flowering Plants" by Alfred Bennett, eschews Goethian rapture in favour of Darwinian sobriety, it is written in the first person (a style that seems to have gone out in the 1950s) and unabashedly uses the epithet 'beautiful'.

By contrast, the tone of the technical contributions to the present issue is doggedly impersonal. Many years of collective effort have gone into the honing of a scientific prose in which an occasional 'we' is sanctioned but in which any personal or collectively enthusiastic voice is systematically repressed.

In this issue, the most prolific users of 'we' appear to be Kulkarni *et al.* on a host galaxy and Schulz *et al.* on climate change — perhaps appropriately in astronomy and geography, in which the role of the observer may still be less readily gainsaid than in many other sciences. More typical in tone are such phrases as "the evidence suggests...", "it is impossible to assess...", "there are only two plausible interpretations..." and "according to the models discussed here..."

**The graphic vehicle is correspondingly clean and business-like, favouring sans-serif typefaces for headers, titles and captions. Its natural visual habitat is the top of a metal desk, beside a computer, or on a modular shelving unit in a modern laboratory.**

This issue has a very different tactile and visual feel from its first predecessor. It is bigger, glossier, thicker, heavier and the range of visual images is far wider. Colour now plays a key role, both decoratively and as a conveyor of information.

All the illustrations in the meagre ration of 11 in 1869 were engraved, two from photographs and one from a spectrometer. The diagram predominated as a conveyor of visual information. The dominant vehicle in this issue is the authoritative graph, in its various forms, of which there are almost 50. Only one image obviously relies on 'drawings' in the traditional sense (page 63), and 'normal' photographs are largely restricted to the News section, Book Reviews and New on the Market. Materials science and genetics help to maintain a decent quotient of sexy visuals,

with the prize for aesthetics going to a reflectance and transmission micrograph of the tip of a diamond (page 47).

Most non-diagrammatic images are generated through the use of devices to detect (and sometimes also to originate) invisible emissions, to do 'perceptual' analysis using computers, and to generate the representations through computer graphics.

The use of machines to perform — in a highly selective way — acts that were formerly the province of human visual perception and hand representation provides the dominant visual tenor of *Nature* in this decade, and perhaps even earlier. Such images breathe an air of non-human objectivity, since our intervention can be minimized during the course of the acts of seeing and showing, but the choices of how the device is designed and operated are no less matters of human judgement than with the hand-drawn image. The key choices now reside at different points in the system and are often less easy to get at for the non-specialist.

Looking at the overall format of the journal, it would be easy to guess that advertisements play an ever greater visual role. However, this guess would be wrong. Not only did the first issue establish the long-term presence of adverts on the cover, but it contained 18 pages of adverts inside compared with 20 of text — to set against the present 36 pages of display adverts, 31 of classified adverts and 99 of text.

The currently fashionable rhetorics of scientific imagery appear in conveniently exaggerated form in the advertisements. Polished photographs of equipment and eager operatives feature, with a lacing of more technical images, such as graphs and the double helix (which must hold the advertisers' record for graphic popularity in this decade). The Eppendorf advert (facing Contents) is particularly telling. Whereas an advert a few years earlier showed the company's pipette with descriptive text, the new Multipette Plus takes second place to an advertising agency's image of thrusting modernity in the form of an ultra-high-speed train, the streamlined profile of which literally outpaces its steam-driven predecessor. Promising to "streamline work in your lab", the Multipette aspires to

"work as reliably and precisely as you yourself" — an oddly inverted compliment, since it will presumably be bought to achieve more precision than is possible with more hand-driven procedures.

The array of advertisements and their disposition reflects a complex series of decisions, based on such factors as historical precedent, willingness to pay premium rates for prime positions, status and leverage, and international sensibilities. Each issue of *Nature* can be seen to embody complex measures of the relative commercial, institutional, personal and national weights that need to be balanced by the editorial and other staff of an international journal of science. Perhaps the only major factors that do not find commensurate visual expression are the huge edifices of the research councils and academic institutions. They tend to feature only through photographs of people and buildings — although, in theory, it should be irrelevant what a major scientist looks like or whether a piece of research originates from an ancient temple or a 1960s greenhouse of glass and steel.

*Nature* in its present guise ensures through its editorial pages that the visual and verbal cultures of scientists play more overt roles than in most specialist journals, yet the dominant feel is of the 'great machine' of science and technology. More traditionally 'subjective' dimensions of the sciences of the natural world — inspirational, passionate, intuitive, poetic, artistic, spiritual, social and political — resurface only as matters of conscious editorial intent, as in the Art and Science series, rather than as part of the integral expression of the substratum of the scientists' mental and material landscapes.

Yet, looking at recent issues of *Nature*, even set beside some of the glossier art periodicals, we are regularly presented with a visual feast which seems to speak of an undiminished aesthetic excitement about acts of observation, analysis, visualizing, modelling and representation. But you would never know from the texts in which the scientists announce their discoveries. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



# Zebra mussels invade Lake Erie muds

Zebra mussels (*Dreissena polymorpha*) originated in western Russia but have now become widespread in Europe and North America. They are widely known for their conspicuous invasion of rocks and other hard substrates in North American and European watersheds<sup>1</sup>. We have found beds of zebra mussels (Fig. 1) directly colonizing sand and mud sediments each year across hundreds of square kilometres of North America's Lake Erie. This transformation of sedimentary habitats into mussel beds represents an unforeseen change in the invasive capacity of this species.

We collected exotic dreissenid mussels and their underlying sediments within 0.25 m<sup>2</sup> quadrats in western Lake Erie from 1994 to 1996. Densities ranged from 1,500 to 32,500 zebra mussels per square metre on sand and sandy-mud sediments. In contrast to previous observations<sup>2,3</sup>, zebra mussel densities in the surveyed sedimentary habitats exceeded those of a co-occurring exotic species, the quagga mussel (*Dreissena bugensis*), by a factor of ten. Independent collections of zebra mussels from other sedimentary habitats in Lake Erie<sup>4</sup> also uncovered densities of more than 20,000 per square metre, demonstrating the widespread establishment of the species in offshore habitats.

These high zebra mussel densities are comparable to those found on hard substrates in North American watersheds following their introduction during the late 1980s (ref. 5) and surpass those reported among well-established European populations<sup>6</sup>. These Lake Erie offshore densities are an order of magnitude higher than those

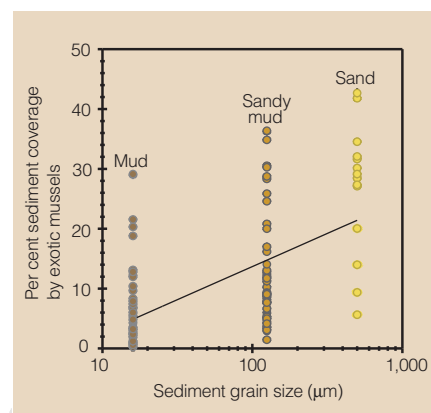
previously identified in sedimentary habitats from either continent<sup>6</sup>.

We found that the sediments underlying the Lake Erie offshore assemblages of zebra mussels lack any associated hard substrates, such as clam shells<sup>7</sup>, which are generally thought to be needed as recruitment nuclei<sup>6</sup>. Moreover, greater than 95% of the underlying sediment (dry weight) had grain sizes of less than 500 µm diameter, with those in the predominant fraction being less than 63 µm. Subsequent microscopic examination of these sediments revealed juvenile zebra mussels (less than 10 mm in length) with byssal threads attached to individual sand grains smaller than 1 mm.

These observations suggest that zebra mussels can directly colonize sand substrates and then use their byssal threads to bind sediments into conglomerates. Subsequent colonization of these sediment-mussel aggregates, enhanced by gregarious recruitment<sup>8</sup>, would provide a positive feedback for the expansion of zebra mussel assemblages in sedimentary habitats.

The continuity and breadth of the range of sizes (3–28 mm) found among the zebra mussels in 1994 indicate that their offshore assemblages were at least three years old at the time. This is confirmed by independent video records, which show that zebra mussels were indeed colonizing sedimentary habitats in Lake Erie by 1991 (ref. 9).

Since the early 1990s, zebra mussels have also been clogging industrial screens associated with commercial sand-dredging in Lake Erie (R. Kight, Erie Sand & Gravel Company, personal communication; April 1997). These observations indicate that zebra mussel



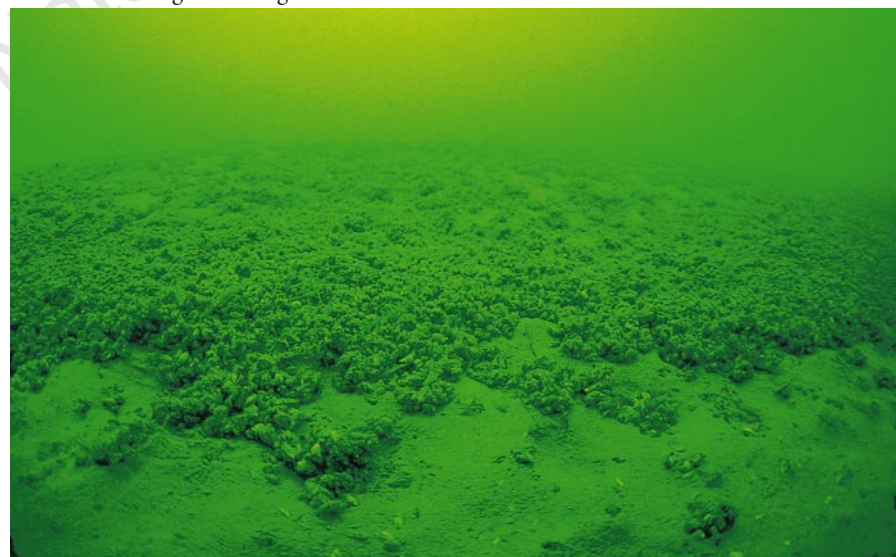
**Figure 2** Relationship between sediment grain size and percentage coverage by exotic dreissenid mussels. Data were derived from side scan sonar surveys<sup>11,12</sup> across known sediment types<sup>10</sup> along a 200-km transect parallel to the coastline at depths of 10–20 m in central and western Lake Erie in 1995 (ref. 13). Digital analyses, using the program Imagine (ERDAS), of 122 random profiles from this transect, each of which covered 25,000 to 50,000 m<sup>2</sup>, indicated that the percentage coverage of invading mussel assemblages in these sedimentary habitats changed significantly with underlying sediment grain size ( $r^2 = 0.347$ , d.f. 120,  $P < 0.001$ ). Estimated means and standard errors of the exotic mussel coverages on Lake Erie sediment types were: sand (500–125 µm)  $26.7 \pm 2.9\%$ ,  $n = 13$ ; sandy mud (125–16 µm)  $12.5 \pm 1.1\%$ ,  $n = 55$ ; and mud (16–4 µm)  $5.8 \pm 0.8\%$ ,  $n = 54$ .

expansion across sedimentary habitats has coincided with their relatively well-studied invasion of hard substrates in Lake Erie.

The coverage of exotic mussel assemblages across known sedimentary habitats<sup>10</sup> was estimated using side scan sonar (SSS), which produces active sonar pulses which are backscattered weakly by sediments and strongly by any overlying hard substrates<sup>11,12</sup>. As in other studies<sup>4,9</sup>, videographic observations<sup>11</sup> were used to ascertain whether hard substrates identified from the SSS surveys were assemblages of exotic mussels.

These SSS analyses were extrapolated (Fig. 2) on the basis of the overall distribution of sediment types (sand, 10%; sandy mud, 22%; and mud, 48%) in this 25,300-km<sup>2</sup> lake ecosystem<sup>12</sup>. From this extrapolation, we estimate that invasive mussel beds were already covering  $2,076 \pm 663$  km<sup>2</sup> (95% confidence interval) of the sediment area in Lake Erie by 1995.

Further investigation is needed into the impact on the ecosystem of these assemblages. The absence of fundamental data on the expansion of such vast zebra mussel beds across sedimentary habitats is a significant shortcoming in understanding the limits and consequences of their invasion, not only in



**Figure 1** Invading zebra mussels (*Dreissena polymorpha*) on fine-grain sand sediments at 14 m depth in western Lake Erie in 1994. Adult mussels (10–30 mm in length) are seen extending 2–3 m across and up to 10 cm above the sediment surface. The presence of erect macrophytes at this depth demonstrates the increased water clarity resulting from extremely high filtration rates by mussels.

Lake Erie, but also in watersheds elsewhere.

**Paul Arthur Berkman, Melissa A. Haltuch, Emily Tichich**

Byrd Polar Research Center, Ohio State University,  
108 Scott Hall, 1090 Carmack Road, Columbus,  
Ohio 43210, USA

e-mail: paul+@osu.edu

**David W. Garton**

Ohio Sea Grant College Program,  
Ohio State University, 1541 Research Center,  
1314 Kinnear Road, Columbus, Ohio 43212, USA

**Gregory W. Kennedy, John E. Gannon**

Biological Resources Division,  
United States Geological Survey, 1451 Green Road,  
Ann Arbor, Michigan 48105, USA

**Scudder D. Mackey, Jonathan A. Fuller, Dale L. Liebenthal**

Division of Geological Survey,  
Ohio Department of Natural Resources,  
1634 Sycamore Line, Sandusky, Ohio 44870, USA

- Nalepa, T. F. & Schloesser, D. W. (eds). *Zebra Mussels: Biology, Impacts and Control* (CRC, Boca Raton, 1993).
- Mills, E. L. et al. *Am. Zool.* **36**, 271–286 (1996).
- Dermott, R. & Munawar, M. *Can. J. Fish. Aquat. Sci.* **50**, 2294–2304 (1993).
- Coakley, J. P., Brown, G. R., Ioannou, S. E. & Charlton, M. N. *Wat. Air Soil Poll.* **99**, 623–632 (1997).
- Griffiths, R. W., Schloesser, D. W., Leach, J. H. & Kovolak, W. P. *Can. J. Fish. Aquat. Sci.* **48**, 1381–1388 (1991).
- Mellina, E. & Rasmussen, J. B. *Can. J. Fish. Aquat. Sci.* **51**, 1024–1036 (1994).
- Schloesser, D. W., Nalepa, T. F. & Mackie, G. L. *Am. Zool.* **36**, 300–310 (1996).
- Ram, J. L., Fong, P. P. & Garton, D. W. *Am. Zool.* **36**, 326–338 (1996).
- McQuest Marine Sciences. *Report on SSS Surveys on Three Sites in Lake Erie: Chicken and Hen Island, East Sister Island and Chickenolee Reef* (Canadian Ministry of Natural Resources, Ontario, 1991).
- Thomas, R. L., Jaquet, J.-M., Kemp, A. L. W. & Lewis, C. F. M. *J. Fish. Res. Bd. Can.* **33**, 385–403 (1976).
- Stanic, S., Briggs, K. B., Fleischer, P., Sawyer, W. B. & Ray, R. I. *J. Acoust. Soc. Am.* **85**, 125–136 (1989).
- Edsall, T. A. et al. *Use of Remote-Sensing Techniques to Survey the Physical Habitat of Large Rivers*, Great Lakes Science Center Contribution No. 983 (US Geological Survey, Ann Arbor, 1997).
- Fuller, J. A. *Distribution of Surficial Sediments in Ohio's Nearshore (Lake Erie) as Interpreted from SSS and 3.5 kHz Sub-Bottom Data*. Open-File 96–520 (US Geological Survey, Reston, 1996).

## Did algal toxins cause monk seal mortality?

The population of Mediterranean monk seals off the coast of the western Sahara has recently suffered a sudden mortality. A morbillivirus was isolated post-mortem from the tissues of three seals<sup>1</sup>, and it has been proposed that the virus was the agent responsible. This conclusion is called into question by epidemiological, clinical, pathological and toxicological considerations. We suggest here that intoxication by algal toxins is a more likely cause of the deaths.

The monk seal, *Monachus monachus*, is a highly endangered species with a worldwide population of fewer than 500 individuals. We examined 117 carcasses from May to July 1997. The mortality affected mainly adults: only 3 juveniles out of 38 in the

colony have so far been found dead. Five pups present at the onset of the event were clinically healthy and still alive at the time of writing, even though four of their mothers had died. This mortality pattern is inconsistent with other recent morbillivirus outbreaks in aquatic mammals, in which juveniles were more severely affected than adults<sup>2,3</sup>.

Most of the seals died at sea, and 14 of the fresher carcasses were analysed. Our observations of these carcasses, together with behavioural observations of terminally ill animals, showed that the seals were in a good nutritional state, with no pulmonary emphysema or gross lesions in evidence.

The seals' lungs were congested and their lungs and airways were filled with fluid. The terminally ill animals also exhibited clinical signs of lethargy, motor incoordination and paralysis in the water: animals floated horizontally, apparently paralysed and incapable of effective voluntary movement, although still breathing and not undergoing convulsions or respiratory distress. The period between the onset of clinical signs and the death of the seals was short. These conditions are consistent with drowning caused by paralysis due to poisoning.

Osterhaus et al., who isolated the morbillivirus in affected seals<sup>1</sup>, presented no histopathological data. However, we carried out histopathological examination of lung and other tissues (S. Kennedy, personal communication) from the 14 fresh carcasses, and found no indication of typical morbillivirus lesions. In particular, there was no evidence of primary viral damage or secondary opportunistic infections in lung tissue, which are hallmarks of morbillivirus infections in other aquatic mammal species<sup>4</sup>.

We analysed eight water samples collected from near the colony during the mortality event, and identified three species of toxic dinoflagellates (*Alexandrium minutum*, *Gymnodinium catenatum* and *Dinophysis acuta*). *A. minutum* was by far the most abundant dinoflagellate (2,000 to 16,500 cells per litre). We successfully isolated and cultured it at the Vigo laboratories. We did not succeed in obtaining cultures from *G. catenatum* and *D. acuta*.

We analysed tissues from eight dead seals, 18 fish collected in the feeding grounds and ten composite bivalve samples, using mouse bioassays and high-performance liquid chromatography (HPLC) to search for paralytic toxins. (Our complete procedures, and results from phylogenetic and toxicological studies, will be published elsewhere.) Mouse bioassays performed by the AOAC<sup>5</sup> method yielded results below and close to the detection limit (30–40 µg saxitoxin (STX) equivalents per 100 g of tissue) in six seal samples and positive results in two fish (*Diplodus sargus* and *Dicentrarchus punctatus*). The

symptoms observed in the mice were consistent with the effects of paralytic toxins, but all the bivalve samples were negative by mouse bioassay.

Using reverse-phase HPLC with post-column reaction and fluorometric detection<sup>6</sup> (detection limit from 0.25 µg for decarbamoyl saxitoxin (dcSTX) to 5.0 µg for neosaxitoxin (NeoSTX) per 100 g tissue), we found variable levels of the PSP toxin dcSTX, NeoSTX and gonyautoxin 1 (GTx1) in all seal samples (liver, kidney, skeletal muscle and brain) and in all fish. Although the toxic profiles of seal and fish were similar, they did not fit with that of the *A. minutum* isolated, suggesting that other organisms were involved. Seal-liver samples ( $N = 7$ ) yielded a mean concentration of  $12.88 \pm 10.09$  s.d. (range 0.2–28.0) µg dcSTX per 100 g of tissue and  $3.48 \pm 2.04$  s.d. (range 0.8–6.4) µg dcSTX per 100 g of tissue in brain samples ( $N = 5$ ).

In bioassay-positive fish, we found 27.6 µg (*D. sargus*) and 90.0 µg (*D. punctatus*) dcSTX per 100 g of viscera. In the remaining 16 fish, including ten different species, we found a mean concentration of  $6.1 \pm 9.14$  s.d. µg dcSTX per 100 g of viscera. The mean dcSTX concentration in bivalve samples was  $3.2 \pm 3.48$  µg dcSTX per 100 g tissue.

This evidence suggests that poisoning by paralytic toxins has played a role in the monk seal deaths. Poisoning with algal toxins has been proposed as a possible aetiology in the mortalities of several marine mammals, including Hawaiian monk seals, humpback whales, manatees, sea otters and even domestic livestock<sup>7–10</sup>. Nevertheless, because information on the dynamics and kinetics of these toxins in marine mammals is scarce, it is impossible to assess whether the levels of toxin we measured were enough to produce death in the Mediterranean monk seals. Mortalities in seabirds or in other marine mammals could not be surveyed in the area.

**Mauro Hernández**

Laboratorio Forense de Vida Silvestre,  
P. Conde de los Gaitanes, 145, La Moraleja,  
28109 Madrid, Spain

**Ian Robinson**

Norfolk Wildlife Hospital, RSPCA, Station Road,  
East Winch, King's Lynn, Norfolk PE32 1NR, UK

**Alex Aguilár**

Department of Animal Biology, Faculty of Biology,  
University of Barcelona, Barcelona 08071, Spain

**Luis Mariano González**

Ministerio de Medio Ambiente,  
Gran Vía San Francisco 4, 28005 Madrid, Spain

**Luis Felipe López-Jurado**

Department of Biology,  
University of Las Palmas de Gran Canaria,  
Canary Islands, Spain

**María Isabel Reyero**

Centro Oceanográfico de Vigo, Instituto Español de  
Oceanografía, Apdo 1552, 36200 Vigo, Spain

**Emiliano Cacho**



Community Reference Laboratory on Marine  
Biotoxins, Ministerio de Sanidad y Consumo,  
Estación Marítima sn, 36271 Vigo, Spain

José Franco

Instituto de Investigaciones Marinas de Vigo,  
Apdo 1552, 36200 Vigo, Spain

Victoria López-Rodas, Eduardo Costas

Department of Genetics, Veterinary Faculty,  
Complutense University, 28040 Madrid, Spain

1. Osterhaus, A. et al. *Nature* **388**, 838–839 (1997).
2. Calzada, N. et al. *Mar. Mamm. Sci.* **10**, 299–310 (1994).
3. Harkonen, T. & Heide-Jorgensen, M. P. *Ophelia* **32**, 211–235 (1990).
4. Kennedy, S. et al. *Veterinary Pathology* **26**, 97–100 (1989).
5. Dunningan, P. J. et al. *Mar. Mamm. Sci.* **12**, 499–515 (1996).
6. Franco, J. M. & Fernandez, P. *Chromatographia* **35**, 613–620 (1993).
7. O'Shea, T. J. *Mammal. Sci.* **7**, 165–179 (1991).
8. Geraci, J. R. et al. *Can. J. Fish. Aquat. Sci.* **46**, 1895–1898 (1989).
9. Gilmartin W. G. et al. in *Proceedings on the Symposium on Status of Resource Investigation in the NW Hawaiian Island, April 24, 1980* (eds Grigg, R. & Fund, R.) 32–41 (Univ. Hawaii, Honolulu, 1980).
10. Witeck, M. et al. *Limnol. & Oceanogr.* **36**, 393–404 (1991).

## ATP synthase's second stalk comes into focus

Almost all the adenosine triphosphate (ATP, the predominant energy-transfer mediator in living organisms) used by cells is made by ATP synthases, which have two parts known as  $F_1$  and  $F_0$ . These enzymes are found in bacterial plasma membranes, chloroplast thylakoid membranes and mitochondrial inner membranes; they function as rotary motors, coupling nucleotide binding in the  $F_1$  part to proton translocation by the transmembrane  $F_0$  part<sup>1,2</sup>. Here we present cryo-electron microscopy images showing details of the connection between the  $F_1$  and  $F_0$  parts of the enzyme.

The bacterium *Escherichia coli* contains

the simplest form of the enzyme, EC  $F_1F_0$ . It comprises eight different subunits, five in the  $F_1$  part ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ) and three in the  $F_0$  part ( $a$ ,  $b$ ,  $c$ ,  $c_{10-12}$ ) (ref. 3). The proton pore is formed at the interface between the single-copy 'a' subunit and the 'c' subunits<sup>4</sup>, which are arranged as a ring<sup>5</sup>.

The  $\gamma$  subunit sits within a ring of alternating and hexagonally arranged  $\alpha$  and  $\beta$  subunits in the  $F_1$  part<sup>6</sup>. It extends from  $F_1$  and, in association with the  $\epsilon$  subunit, forms the 40–45-Å-long stalk that has been seen by electron microscopy to link the  $F_1$  and  $F_0$  parts<sup>7</sup>. Both the  $\gamma$  and  $\epsilon$  subunits bind to the c subunit ring, and the emerging mechanism of energy coupling<sup>1,8</sup> involves the  $\gamma$ ,  $\epsilon$  and c subunit ring rotating as a unit relative to the fixed  $\alpha_3\beta_3$  domain of the  $F_1$  and  $ab_2$  domain of  $F_0$ .

Biochemical evidence supports the idea of a second connection<sup>9</sup> between  $F_1$  and  $F_0$ , but this feature had not been observed directly. Previous electron microscopy studies used membranous EC  $F_1F_0$  embedded in a thin layer of ice. These studies could have missed a narrow second stalk, because the density difference between protein and solvent water is low and consequently the images are noisy, so many must be averaged to see any features clearly. Also, the membranous nature of the sample often resulted in superimposition of individual complexes in projection, making interpretation of weak features more difficult.

We used detergent-solubilized EC  $F_1F_0$ , prepared and examined after negative staining to enhance the contrast between protein and solvent. Staining with heavy metals had previously been avoided because this treatment led to release of  $F_1$  from  $F_0$ ; however, we found that reaction of the enzyme with dicyclohexylcarbodiimide helped to prevent this release.

We observed two stalks linking the  $F_1$  to

the  $F_0$  (Fig. 1) in about 40% of all images obtained. The fatter, more central stalk, which had been observed previously, contains the  $\gamma$  and  $\epsilon$  subunits. The second connector is at the side of the molecule, clearly evident extending down from the  $F_1$ , and a corresponding density rises up from the  $F_0$ . This second stalk probably includes both the  $\delta$  and the b subunits. Density in the middle of the stalk is weak, as would be expected if this region derives from one or two  $\alpha$ -helices from each of the b subunits.

At the top of the  $F_1$  is a cap that extends in the direction of the more asymmetrically placed stalk. Such a feature is lacking in the X-ray structure reported for bovine  $F_1$  (ref. 6). It is probable that this cap is formed by the amino-terminal 30 or so residues of the  $\alpha$  subunits — which were unresolved in the X-ray experiments — together with a part of the  $\delta$  subunit.

There is also a clear asymmetry of the  $F_0$  part, consistent with a substructure in which the c subunits form a ring with the a and b subunits outside<sup>5,10</sup>. This asymmetry has been observed in enzyme solubilized with amphipol (Fig. 1), as well as in lysolipid and laurylmaltoside (results not shown). So, although detergent binding to the  $F_0$  part may contribute to the effect, it is likely that the observed asymmetry is a feature of the protein.

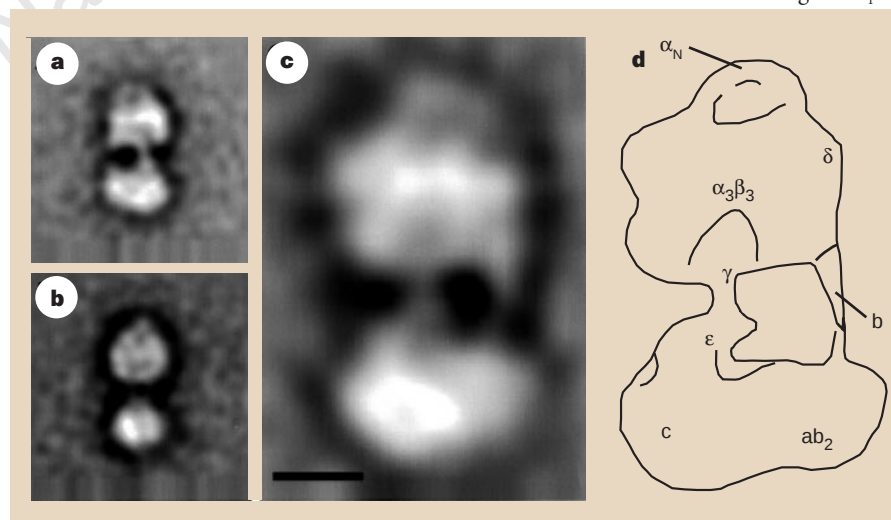
The presence of the second stalk — which is also evident in electron micrographs of the less well-defined  $V_1V_0$ -type ATPase from another bacterium, *Clostridium fervidus*<sup>11</sup> — has important functional implications. It could provide the stator against which the  $\gamma$ - $\epsilon$  subunit crankshaft rotates<sup>1</sup>. Thus, ATP hydrolysis in one direction and ATP synthesis in the other would rotate the  $\gamma$ - $\epsilon$ -c subunit domain, thereby mechanically coupling nucleotide binding in catalytic sites with proton translocation.

Stephan Wilkens\*, Roderick A. Capaldi

Institute of Molecular Biology, University of Oregon,  
Eugene, Oregon 97403-1229, USA

e-mail: rcapaldi@oregon.uoregon.edu

\*Present address: Department of Biochemistry,  
University of California at Riverside, Riverside,  
California 92507, USA



**Figure 1** Electron microscopy images of ECF<sub>1</sub>F<sub>0</sub> in side view. **a–c**, Different views based on image analysis and classification of a data set of 139 single molecules; **a** is a mirror image of **c**; **b** appears to be a projection at around 60° to that in **a** and **c**. **d**, Our interpretation of the arrangement and composition of the second stalk, and the extra density on top of the  $F_1$  molecule.

1. Junge, W., Lill, H. & Engelbrecht, S. *Trends Biochem. Sci.* **22**, 411–452 (1997).
2. Noji, H., Yasuda, R., Yoshida, M. & Kinosita, K. Jr *Nature* **386**, 299–302 (1997).
3. Weber, J. & Senior, A. E. *Biochim. Biophys. Acta* **1319**, 19–58 (1997).
4. Fillingame, R. H. *Curr. Opin. Struct. Biol.* **6**, 491–498 (1996).
5. Singh, S., Turina, P., Bustamante, C.J., Keller, D.J. & Capaldi, R. A. *FEBS Lett.* **397**, 30–34 (1996).
6. Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. *Nature* **370**, 621–628 (1994).
7. Gogol, E. P., Lücken, U. & Capaldi, R. A. *FEBS Lett.* **219**, 274–278 (1987).
8. Boyer, P. D. *Annu. Rev. Biochem.* **66**, 717–749 (1997).
9. Ogilvie, I., Aggeler, R. & Capaldi, R. A. *J. Biol. Chem.* **272**, 16652–16656 (1997).
10. Birkenhäger, R., Hoppert, M., Deckers-Heberstreit, G., Mayer, F. & Altendorf, K. *Eur. J. Biochem.* **230**, 58–67 (1993).
11. Boekema, E. J., Ubbink-Kok, T., Lolkema, J. S., Brisson, A. & Konings, W. N. *Proc. Natl Acad. Sci. USA* **94**, 14291–14293 (1997).



# All aboard the biotech express

## The Biotech Century: Harnessing the Gene and Remaking the World

by Jeremy Rifkin

Tarcher: 1998. Pp. 259. \$24.95

Sheldon Krinsky

This year marks the twenty-fifth anniversary of the discovery of recombinant DNA technology, or gene splicing. Various symposia and publications have paid tribute to the pioneers of genetic engineering and the new industrial developments it has spawned. The largely publicly funded field of molecular biology has provided tools, knowledge and product ideas that have attracted substantial investments from the world's pharmaceutical, diagnostics and agricultural sectors. Designer drugs, genetic screening tests and transgenic crops are moving rapidly from the Petri dish to the marketplace. But throughout this period, critics have questioned the uses to which genetic technology is being put. Of all the sceptics, Jeremy Rifkin has been the most visible in the media and the most maligned by the principal players in the new bio-academic-industrial complex.

*The Biotech Century* is Rifkin's third book devoted primarily to biotechnology. In 1977 he was a co-author of *Who Shall Play God?*, which describes the rise of a new eugenics movement and warns the reader that "it would be wrong to proceed with genetic engineering now". A second work, *Algeny*, published in 1983, is a selective compendium of footnotes in the history of science that positions biotechnology within a neo-Darwinist *Weltgeist* and posits two paths to the future, one with genetic engineering and one without it. ("There could be no lonelier place than a biologically engineered world.")

In *The Biotech Century*, Rifkin is no less passionate in his opposition to a genetically engineered world, but he makes a symbolic effort to dissociate himself from his image as biotechnology's "consummate contrarian", which he skilfully helped to create. He says: "In fact, there is value,

great value, in some of the products of genetic engineering and that's what makes the discussion of this ultimate human technology so interesting, difficult, and challenging." This is the one moment in the book where we are led to anticipate a more complex analysis of the many sides to biotechnology. We discover soon enough that it is an empty concession. Rifkin does his best work in drawing attention to the growing inventory of real and potential dangers and the ethical conundrums raised by genetic technologies. For this role he is admired by those disenfranchised from technological choices who see him as their *vox populi*.

This book is in large measure a response to the realization of Francis Bacon's *The New Atlantis*, published in 1622, in which Bacon describes a future science that uses the existing forms of biological life as the raw materials from which to redesign nature. Rifkin's chapters "A second genesis", "Reinventing nature" and "A eugenic civilization", reflecting the Baconian prophecy, raise serious questions. Can we redesign living things to work better for us while protecting the integrity of our natural heritage? Who is the 'us' on whose behalf biotechnology is being developed? Will herbicide-tolerant crops yield a more nutritious and plentiful bounty of food that will be priced and distributed

fairly? Does milk production using synthetic growth hormone meet the public need or solve the problem of rapidly declining small farms? Will those who lost out in the genetic lottery be helped by the new generation of genetic tests or will they suffer discrimination for their 'pre-existing condition'?

Some will undoubtedly read this work as another anti-science polemic, but they will not have read it correctly. It is too equivocal to draw such a conclusion. Rifkin shows signs of acknowledging that the 'biotechnology express' has long since left the station. After reports of genetic advances that could serve as copy for a promotional brochure on biotechnology, the early chapters present an inventory of worthy cautions.

The chapter on patenting takes us from the 1980 US Supreme Court decision that set the broad legal mandate for patenting living things to current trends in patenting genes, indigenous crops and human cell lines. Rifkin himself, seeking to test the public's tolerance for the bizarre implications of patenting life, recently teamed up with a scientist and applied for a patent on a human-animal chimaera (see *Nature* 392, 423; 1998), a concept he raises in the book. Other chapters are rich in examples but deficient in the more complex analysis of how largely decentralized and diverse industrial

## Ornithological opus

The imperial eagle (*Aquila heliaca*), from a plate in *The Birds of the Western Palearctic* edited by D. W. Snow and C. M. Perrins (Oxford University Press, £150). At more than 1,600 pages comprising two volumes, this is the "concise" updated edition of a nine-volume handbook published between 1977 and 1994.



sectors producing new technologies and products can be properly managed and publicly overseen to avoid ecological, ethical or human-health mishaps.

Some of this complexity is acknowledged in the final chapter, "A personal note", in which Rifkin writes of using science, and even genetics, in a manner that respects our natural world: "the question is what kind of biotechnologies will we choose in the coming Biotech Century?" Our greatest challenges lie in the social guidance and assessment of biotechnology within a democratic participatory framework and a global awareness.

Rifkin was criticized 20 years ago for exaggerating the untoward paths biotechnology would take and for opposing scientific progress. In hindsight, many of his predictions can hardly be considered hyperbole. Serious discussions are taking place on cloning humans, altering human germ cells, universal genetic screening, mandatory DNA identification and even the unmentionable prospect of 'improving' the human gene pool.

In his role as social critic of biotechnology, Rifkin has become entangled in a paradoxical situation. In an attempt to stir the reader's passions, he overdramatizes in some sections the power of biotechnology: "The biotechnology revolution will affect each of us more directly, forcefully, and intimately than any other technology revolution in history." We must be reminded that the first three main agricultural products of genetic engineering — the Flavr Savr tomato, the Ice Minus bacterium and recombinant bovine growth hormone — have either failed or are failing.

Moreover, like many of the genetic scientists he calls to task, Rifkin at times accepts uncritically a view that vastly overestimates the importance of genes in biological organisms. ("With genetic engineering, we assume control over the hereditary blueprint of life itself.") If he were to question too seriously the ability of science to carry out its "redesign of nature", the book would lose much of its moral force. But, while he plays up the power of genes ("the ultimate exercise of power") in some sections, elsewhere he tempers that view by acknowledging the poverty of genetic reductionism. He points out that, although such a view may be false, it has helped to advance the interests of those involved in molecular biology.

There is no simple reading of this book. A fair-minded reader will agree with some points and disagree with others. At a time when scientific institutions are struggling with the public understanding of science, there is much they can learn from Rifkin's success as a public communicator of scientific and technological trends. □

Sheldon Krinsky is in the Department of Urban and Environmental Policy, Tufts University, 97 Talbot Avenue, Medford, Massachusetts 02155, USA.

## Living with dinosaurs

### The Rise of Birds: 225 Million Years of Evolution

by Sankar Chatterjee  
Johns Hopkins University Press: 1998.  
Pp. 312. \$39.95, £33

### The Mistaken Extinction: Dinosaur Evolution and the Origin of Birds

by Lowell Dingus and Timothy Rowe  
W. H. Freeman: 1997. Pp. 332. \$34.95, £24.95

### Taking Wing: Archaeopteryx and the Evolution of Bird Flight

by Pat Shipman  
Simon and Schuster: 1998. Pp. 336. \$25. To be published in the United Kingdom in June by Weidenfeld and Nicolson, £20

José Luis Sanz, Bernardino P. Pérez-Moreno and Francisco J. Poyato-Ariza

In recent years there has been a renaissance in Mesozoic palaeo-ornithology. This kind of reinvigoration is normal in palaeontology, as new discoveries show that old hypotheses are false and give rise to new ones. A large number of Mesozoic avian taxa have been discovered during the past 20 years, so it is perhaps not surprising that three books devoted to Mesozoic palaeo-ornithology have recently been published. They are different in structure and scope, but the main topics of each are the foremost areas of research on the early evolutionary history of birds: the origin and historical diversity of the avian clade, its phylogenetic relationships, and the development of flight.

One of the most stimulating ideas arising from vertebrate palaeontology is the dinosaurian origin of birds. Proposed by Thomas Huxley in 1868, and reformulated by John Ostrom in the 1970s, this hypothesis is now widely accepted, and is supported by a large amount of evidence that increases as new fossil forms are discovered. All three books favour this hypothesis, which is challenged nowadays by just a few researchers. In terms of current phylogenetic systematics, the dinosaurian origin hypothesis implies that birds have to be considered as short-tailed, feathered volant dinosaurs.

The interpretation of birds as present-day dinosaurs leads to the provocative idea that dinosaurs have been around ever since humans appeared on Earth, and therefore represent an important part of our natural, cultural and economic environment. So it is not surprising that palaeontologists are concerned by man's responsibility in the disappearance of extant dinosaurs, given that natural phenomena — the Cretaceous/Tertiary (K/T) biotic crisis — extinguished most of the diversity of Upper Cretaceous dinosaurs (including some avian ones, such as Enantiornithes, Hesperornithiformes and Ichthyornithiformes).

Sankar Chatterjee joins Lowell Dingus and Timothy Rowe in adopting a militant point of view, describing how the pressures applied by man have led to the extinction of hundreds of modern dinosaur species. The causes of the K/T extinction are still under debate, but it is clear that man has caused many extinctions in the Holocene.

Chatterjee concludes that the impact hypothesis was the proximate cause of the K/T biotic crisis, while volcanic phenomena increased the climatic stress and enhanced the extinction process. Dingus and Rowe review earlier hypotheses about the K/T biotic crisis, and then contrast the volcanic and impact hypotheses and revise the patterns of extinction and survival. A clear consensus has yet to emerge, but many present-day researchers consider that the extinction of non-avian dinosaurs was caused by a combination of the volcanic/marine regression and impact hypotheses. So large extraterrestrial impacts and massive eruptions of flood basalts have to be considered.

The historical diversity of a group of living organisms is shaped by extinction and diversification, the pattern of which is mapped by a phylogenetic hypothesis. The discovery of new fossils shows that the phylogenetic history of Mesozoic birds is much more complex than previously thought. Chatterjee agrees with Dingus and Rowe — and most palaeo-ornithologists — that the phylogenetic map of birds is shaped by a series of successive sister taxa from *Archaeopteryx* to Neornithes (extant birds). Like that of Dingus and Rowe, the avian phylogenetic hypothesis put forward by Pat Shipman is within the consensus reached by most researchers.

Chatterjee, however, includes a problematic taxon, the Triassic *Protoavis*, which is the core of his book. He positions this enigmatic genus between *Archaeopteryx* and the Enantiornithes, implying that *Protoavis* is a basal bird but is more derived than *Archaeopteryx*. The combination of characters of *Protoavis* is very unusual because, according to Chatterjee's interpretation, it has an ornithothoracic-like pectoral region associated with a very primitive (basal archosaur-like) hand architecture. This combination challenges the transformation sequence of characters in early avian evolutionary history.

If *Protoavis* was a bird (between 60 million and 70 million years older than *Archaeopteryx*), this hypothesis predicts important modifications even in the evolutionary history of non-avian dinosaurs. A problematic consequence is that even derived theropod groups, such as dromaeosaurids, troodontids or tyrannosaurids, must appear during the Lower or Middle Triassic. Chatterjee's answer to this troubling idea is highly classical: imperfection and sampling scarcity in the fossil record. Nevertheless, that tyrannosaurids



are about the same age or even older than the first known dinosaurs, such as *Eoraptor* or the herrerasaurids, seems improbable.

So where does *Protoavis* fit in? It could be a taphonomic assemblage of different animals. But there are some bony elements that clearly exhibit bird-like characters. The *Protoavis* evidence could also suggest either the presence in the Upper Triassic of a non-avian theropod or other archosaur with convergent bird traits, or the appearance of the avian clade more than 200 million years ago. Present-day evidence is much more consistent with the first hypothesis.

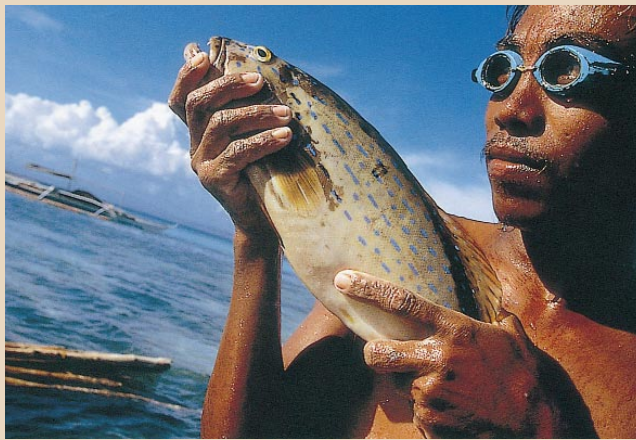
Although all three books fit in with the consensus that birds are flying dinosaurs, there are differences in their proposals about the origin of flight. There are two hypotheses to account for the origin of flight: "from the trees down" (usually associated with a non-dinosaurian ancestor of birds), and "from the ground up" (usually associated with a dinosaurian ancestor of birds). Shipman agrees with Dingus and Rowe in maintaining the consensus on the origin of flight, following the "from the ground up" hypothesis, suggesting that avian flight originated from a cursorial biped (a theropod dinosaur), whereas Chatterjee suggests a new synthesis, proposing a dinosaurian ancestor of birds and a "from the trees down" origin of flight. He considers some theropod clades, commonly thought to be highly cursorial animals, as arboreal ones. This requires new functional interpretations of some structures: for example, the stiffened caudal appendage must be reinterpreted as a climbing prop. Nevertheless, there are no features that clearly indicate an arboreal habitat for theropods such as ornithomimids, whereas nearly every feature indicates that they are extremely well adapted for high-speed running.

Shipman makes an outstanding analysis of the contribution of *Archaeopteryx* to discussions about the origin of flight. She considers the subject from the impartial point of view of someone from another research field. This allows her, after analysing all the hypotheses, to reach a conclusion based mainly on a functional approach: that birds come from a cursorial theropod ancestor.

Shipman's book is an accessible, yet precise, account. It provides a solid background to the importance of *Archaeopteryx* for understanding the origin of avian flight, and is suitable for specialists and non-specialists. Dingus and Rowe focus on dinosaurs, dealing with birds and the origin of flight within the context of the K/T boundary extinctions. They present much information, provide a good phylogenetic background and use familiar language. They construct a hypothetical, conceptual and methodological framework that could be assumed by a large part of the scientific community. Chatterjee takes a more specialist point of view. The other two books present a comprehensive

## Fishing for solutions

"While fisheries around the world are in trouble, it's not too late to turn the tide," we read in *Faces of Fishing: People, Food and the Sea at the Beginning of the 21st Century* (Monterey Bay Aquarium, \$19.95). Bradford Matsen presents both the problems and potential remedies, in a book based on images collected for an exhibition held by the Californian aquarium.



survey of the most relevant alternative hypotheses, but Chatterjee's is different because most of the main hypotheses are his own and are not widely accepted. □

José Luis Sanz, Bernardino P. Pérez-Moreno and Francisco J. Poyato-Ariza are in the Unidad de Paleontología, Departamento de Biología, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain.

## Farming forecast

**Climate Change and the Global Harvest: Potential Impacts of the Greenhouse Effect on Agriculture** by Cynthia Rosenzweig and Daniel Hillel  
Oxford University Press: 1998. Pp. 324.  
£49.50, \$65

Peter D. Moore

The public expects science to be predictive, but most scientists are sufficiently aware of their own ignorance to avoid prediction wherever possible and to qualify it with wide error margins if speculation is unavoidable. Agricultural scientists are among those of whom most is demanded by politicians and the public, yet they face some of the greatest problems in prediction, as they deal with several levels of compounded errors.

They inherit the innate uncertainties of climate prediction and also have to scale up to global levels from small-scale experiments on the responses of crop plants to simulated sets of conditions. They then need to account for the simultaneous responses of weeds, pests and diseases to the new conditions before they can assess the outcome of change on global crop yields. As the geographical pattern of agricultural practice inevitably changes, there are also matters such as soil conservation, wildlife habitat protection and socioeconomic factors to consider, all of which differ from one area to another. Considering all of these variables together, it is a brave author who is willing to write a book about climate change and the global harvest.

A commendable feature of the book by Cynthia Rosenzweig and Daniel Hillel is their readiness to face the complex factors that will influence the agriculture of the future. Consider, for example, the consequences of altered attitudes to energy production resulting from efforts to avoid carbon emissions to the atmosphere. Energy-expensive agrochemicals will be avoided, reduced tillage agriculture will be demanded, biomass energy plantations may compete with food production, extra land may be flooded for hydro-development, and so on. Rosenzweig and Hillel take great pains to document the involvement of agricultural processes in the emission and control of greenhouse gases into the atmosphere and the consequent feedback effects of agricultural change.

The response of a crop to environmental change can be investigated in several ways. The use of controlled environment chambers has many advantages, such as precise factor control, but may not adequately simulate conditions of soil evaporation. Field experiments using open-topped enclosures have been more satisfactory, but even this approach has disadvantages, such as the difficulty in investigating the interactive effect of increased carbon dioxide at different temperatures. There is also the problem of scaling up from such data in a world of varied microclimate and soil conditions. Such soil variations may mean that the potential advantages of higher carbon dioxide levels for plant growth will not occur in many regions because of soil nutrient limitation.

Using recent experience of atmospheric and agricultural change to test the validity of models appears superficially attractive, and one can calculate the expected increase in global agricultural yield as a result of increasing atmospheric carbon dioxide levels from 280 to 360 p.p.m.v. (the observed change over the past 200 years) at 7.5%. The problem is that improved varieties of crops and enhanced input of nitrogen to soils from atmospheric pollution have also influenced



yields, so the crop response model cannot easily be tested from historical documentation.

Changes in pest populations further complicate the picture, as increased global temperatures will generally lead to higher levels of pest infestation in agriculture. Migratory pests that currently survive the winter in lower-latitude refugia, such as the sunflower moth and the corn earworm, may benefit from extended winter ranges and possibly also from easier transportation from one region to another with new atmospheric circulation patterns.

As well as considering the total global yield of agricultural production, Rosenzweig and Hillel investigate regional effects in relation to local environmental and economic conditions. Modelling suggests that the greatest potential increases in agricultural production will occur in Canada and Russia, with the strongest negative trends in equatorial regions, particularly Central America and South-East Asia. But the outcome is dependent on the extent of the temperature rise and, if the increase in global temperature reaches 4 °C, the impact could be negative throughout the world.

The authors use Egypt and the United States as case studies to examine the likely impact of global change on agriculture in specific areas. They demonstrate the flexibility of the United States in responding to new conditions and regard its agriculture as relatively robust but perhaps sensitive to the anticipated variability of a future climate. Egypt, which depends on irrigation for its agriculture, should be regarded as at risk. A rise in sea level would cause greater erosion and salination of the valuable lands of the Nile delta, and a rise in temperature and evaporation can have only negative effects. Much will depend on Egypt's efficiency in the use of its agricultural land and its limited water resources.

The authors conclude that uncertainties abound. They are cautious in their predictions, but their book must rank as the most integrated, informative and accessible account of the complexities of this subject. □

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

## Stress points

### A Mood Apart: A Thinker's Guide to Emotion and its Disorders

by Peter C. Whybrow  
HarperCollins/Picador: 1998. Pp. 340.  
\$14 (pbk), £17.99 (hbk)

Stuart Sutherland

As anyone working in the field will tell you, our knowledge of the neural correlates of manic depression has vastly increased in recent years. We now have at least 20 drugs

that alleviate depression and several that both reduce mania and have a prophylactic effect on bipolar manic depression. We know — or think we know — which neurotransmitters are involved in affective illness. We also know much about the hormonal and neurotransmitter responses to stress, which are often, although by no means always, a precipitating factor in mental illness. And we have developed a psychological treatment for depression that actually works.

But despite these advances, a great deal remains opaque. The prototypes of most of the drugs were discovered by chance. We have no idea of the detailed neural pathways governing mood. We are uncertain of the aberrations in transmitter function that result in depression or mania. And we do not understand why stress should sometimes be beneficial, sometimes disastrous.

In his account of affective disorders, Peter C. Whybrow emphasizes the importance of stress. He argues that the physical threats with which the brain has evolved to cope were transient, whereas stress induced by social factors can be long-lasting and hence damaging. He repeatedly claims that homeostasis is disorganized by stress. True, the brain contains an incredible number of homeostatic mechanisms but, unless the way these mechanisms are affected can be specified, the use of the term "homeostasis" serves only to conceal our ignorance.

Although Whybrow knows his way around the brain's biochemistry, he is sometimes rather dogmatic. Perhaps he thought it would tax the reader too heavily if he mentioned explanations antithetic to his own. For example, he believes that depression is in part caused by underactivity in the monoamine systems, particularly that of dopamine, but many believe that overactivity is the cause. The drugs that ameliorate depression take several weeks to work and their efficacy may depend on a reduction of post-synaptic receptors in response to a sharp increase in amine activity. Moreover, Whybrow is sometimes cavalier with facts. He writes "Prozac affects only serotonin": in reality, it affects several neurotransmitters, and is the least selective of all the serotonin-specific reuptake inhibitors.

Unfortunately, the book goes more seriously awry when it deals with the psychological aspects of mood. The account of cognitive therapy is unsatisfactory. Whybrow believes that depression is associated with an "inaccurate" view of the world and of oneself, a view that is gloomier than is justified. In reality, the depressive's views are more accurate than those of normal people. The fact is — perhaps not surprisingly in this vale of tears — that people can be happy only if they are unrealistically optimistic. Again, he asserts that losing a mother or having a disturbed childhood can lead to depression in later life, but this is at best contentious.

Indeed, one authority recently wrote that the influence of child abuse on adult personality was "barely detectable".

The book has a rather ponderous style. For example, Whybrow writes of the action potential: "This efficient architectural innovation is essential to the increased speed of signalling that is a key element of survival behaviour", a statement that could have been made using half the number of words.

Whybrow follows current fashion by larding the text with anecdotes about himself, his friends and his patients. The inclusion of such largely irrelevant material is an insult to readers, as it assumes they cannot concentrate on a serious exposition without being constantly diverted by the human touch.

Having been rather critical, I should confess an interest: I am myself about to publish a book on mental illness, some of which overlaps with Whybrow's. I approached his book with the highest expectations, based on his eminence, and maybe I have overreacted to my disappointment; other readers may appreciate it more. Whybrow is extremely learned, as shown by the 70 pages of references and notes, which — at least for me — made considerably better reading than the text. Moreover, from time to time he makes some good points, for example by insisting that depression is not akin to sadness. And those of a more literary disposition may well enjoy the mixture of fact, speculation and emotional tales, some of which might make moving short stories. □

Stuart Sutherland is at the Laboratory of Experimental Psychology, University of Sussex, Falmer, Brighton BN1 9QG, UK.

### New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 10 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1996 and issued at least four separate numbers by the end of May 1998.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language is English.
- Deadline for submission is 5 June.

Please send at least four different issues (the first, the most recent and any two others) of each eligible title, together with full details of subscription rates, to: Peter Tallack, *Nature*, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567. e-mail: p.tallack@nature.com

# Identification of a host galaxy at redshift $z = 3.42$ for the $\gamma$ -ray burst of 14 December 1997

S. R. Kulkarni\*, S. G. Djorgovski\*, A. N. Ramaprakash\*\*†, R. Goodrich‡, J. S. Bloom\*, K. L. Adelberger\*, T. Kundic\*, L. Lubin\*, D. A. Frail§, F. Frontera||#, M. Feroci¶, L. Nicastro\*, A. J. Barth\*\*, M. Davis\*\*, A. V. Filippenko\*\* & J. Newman\*\*

\* Palomar Observatory 105-24, California Institute of Technology, Pasadena, California 91125, USA

† Inter-University Centre for Astronomy and Astrophysics, Ganeshkhind, Pune 411 007, India

‡ W. M. Keck Observatory, 65-0120 Mamalahoa Highway, Kamuela, Hawaii 96743, USA

§ National Radio Astronomy Observatory, Socorro, New Mexico 8801, USA

|| Istituto Tecnologie Studio delle Radiazioni Extraterrestri, CNR, via Gobetti 101, Bologna I-40129, Italy

# Dipartimento di Fisica, Università Ferrara, Via Paradiso 12, I-44100, Italy

¶ Istituto di Astrofisica Spaziale, CNR, via Fosso del Cavaliere, Roma I-00133, Italy

\* Istituto Fisica Cosmica App. Info., CNR, via U. La Malfa 153, Palermo I-90146, Italy

\*\* Department of Astronomy, University of California, Berkeley, California 94720, USA

**Knowledge of the properties of  $\gamma$ -ray bursts has increased substantially following recent detections of counterparts at X-ray, optical and radio wavelengths. But the nature of the underlying physical mechanism that powers these sources remains unclear. In this context, an important question is the total energy in the burst, for which an accurate estimate of the distance is required. Possible host galaxies have been identified for the first two optical counterparts discovered, and a lower limit obtained for the redshift of one of them, indicating that the bursts lie at cosmological distances. A host galaxy of the third optically detected burst has now been identified and its redshift determined to be  $z = 3.42$ . When combined with the measured flux of  $\gamma$ -rays from the burst, this large redshift implies an energy of  $3 \times 10^{53}$  erg in the  $\gamma$ -rays alone, if the emission is isotropic. This is much larger than the energies hitherto considered, and it poses a challenge for theoretical models of the bursts.**

Ever since their discovery nearly three decades ago<sup>1</sup>, it was understood that progress in solving the puzzle of  $\gamma$ -ray bursts (GRBs) depends on their identification at other—preferably optical—wavelengths, so that the distances could be measured using standard spectroscopic techniques. From distances and flux measurements one can then infer luminosities and other physical parameters, which can then be used to test theoretical models of the bursts and their origins.

A recent breakthrough in this field was the precise localization of bursts by the BeppoSAX satellite<sup>2</sup>, which has led to the first identifications of GRBs at other wavelengths: X-rays<sup>3</sup>, optical<sup>4</sup> and radio<sup>5</sup>. This has further led to the determination of the distance scale of GRBs, with the detection of intergalactic absorption lines<sup>6</sup> in the optical transient<sup>7,8</sup> (OT) of GRB970508 (refs 9, 10). Apparent host galaxies have been detected for the first two optical afterglows found<sup>11–14</sup>.

Here we report follow-up studies of the OT<sup>15</sup> of a relatively bright burst, GRB971214 (refs 16–18). As the OT faded away, we found an extended object with a red-band magnitude  $R = 25.6 \pm 0.15$  at the position of the OT. Based on the excellent positional coincidence,  $0.06 \pm 0.06$  arcsec, we argue here that this is the host galaxy of GRB971214. Spectroscopic observations show that the host is a typical star-forming galaxy<sup>19–21</sup> at a redshift  $z = 3.418$ .

Given this high redshift, the  $\gamma$ -ray energy release of this burst is unexpectedly large, about  $3 \times 10^{53}$  erg, assuming isotropic emission, corresponding to about 16% of the rest-mass energy of our Sun. Energy released in other forms of radiation, for example, neutrinos or gravity waves, is not included in this energy budget. Nonetheless, the inferred energy release in  $\gamma$ -rays alone is so substantial that it may present difficulties for some of the currently popular theoretical models for the origin of the bursts (coalescence of neutron stars). We may be forced to consider even more energetic

possibilities<sup>22,23</sup> or to find ways of extracting more electromagnetic energy in coalescence models.

## The optical transient

Our follow-up at the Keck Observatory began on the night of 1997 December 15 UT, approximately 13 h after the burst. The only available imager was a 'guide' camera with a small field-of-view and relatively low sensitivity; consequently, these data were not used in the analysis described below. For the subsequent nights we used the Low Resolution Imaging Spectrograph<sup>24</sup> (LRIS) mounted at the Cassegrain focus of the Keck II 10-m telescope. Our early observations were made in the I band. This choice was driven by the presence of a bright Moon in the sky. Later observations were done under darker sky conditions and with a R-band filter (OG570 + KG3). In Table 1 we give a summary of our observations as well as other reported measurements<sup>25–29</sup>.

Figure 1 shows an I-band image obtained on 1997 December 16.52 UT, when the OT<sup>26</sup> was still bright. I-band and R-band magnitudes of the OT are given in Table 1 and shown in Fig. 2. In order to establish consistency between our measurements and the earlier measurements reported by others, we measured instrumental magnitudes of a number of 'secondary' stars in the general vicinity of the OT and then used them to define the zero-point of our instrumental magnitudes for all of our data sets. We calibrated the instrumental magnitudes by using observations of the standard star PG0231 + 051 (for which we assume I-band magnitude  $I = 16.639$  mag; ref. 30) obtained on 1997 December 22 UT. We then derived the magnitudes for stars H1 and H2 (identified in Fig. 1) as  $16.1 \pm 0.1$  mag and  $18.6 \pm 0.2$  mag, respectively. For these two stars, Halpern *et al.*<sup>15</sup> obtained 15.93 mag and 18.46 mag. Within errors, these two sets of measurements agree and thus allow us to link our I-band magnitudes to those of Halpern *et al.*<sup>26</sup>. The zero

point for our R-band magnitudes was set by assuming  $R = 20.1 \pm 0.1$  for object 2 of Henden *et al.*<sup>29</sup>; see also Fig. 1. Our R-band magnitude is in excellent agreement with that obtained by Diercks *et al.*<sup>28</sup> at about the same epoch.

The decay of previous OTs associated with GRBs appears to be well characterized by a power law:  $S(t) \propto t^{-a}$ , where  $S(t)$  is the flux of the OT and  $t$  is the time since the  $\gamma$ -ray burst. Accepting this parametric form and considering only the time interval December 15–17, we obtain  $I(t) = (22.22 \pm 0.18) + (3.09 \pm 0.52) \log(t)$ ; here  $t$  is in days. As magnitudes are defined to be  $-2.5 \log(S) + \text{constant}$ , the value of  $a$  is  $1.22 \pm 0.2$ —remarkably similar to those reported for previous OTs<sup>31–33</sup>.

From the data in Fig. 2 we see that initially the R-band flux appears to decrease in approximately the same fashion as the I-band flux. Such broad-band decay has been noted in previous OTs<sup>32,33</sup> and indeed is an expectation of simple fireball models<sup>31,34</sup>. Accepting that all optical bands decay with the same  $a$ , we obtain  $R(t) = (23.0 \pm 0.22) + (3.09 \pm 0.52) \log(t)$ . In arriving at this equation we have used our 16 December R-band measurement (Table 1) to set the zero point.

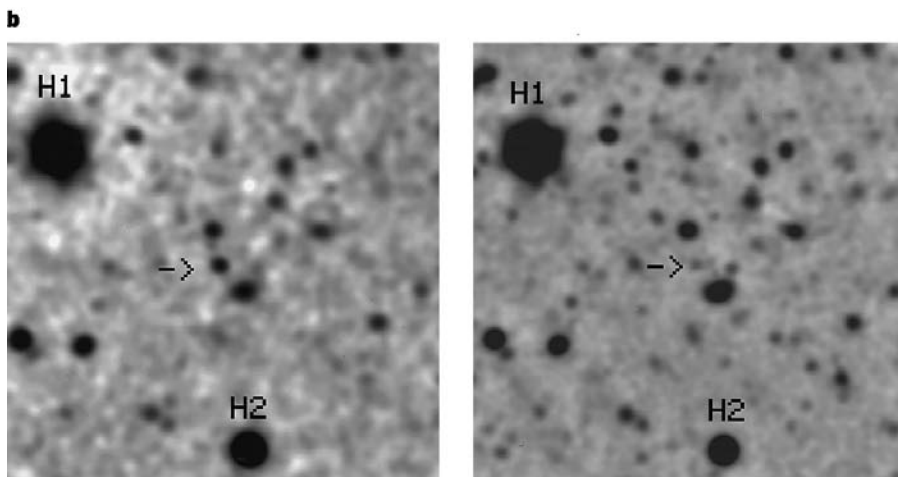
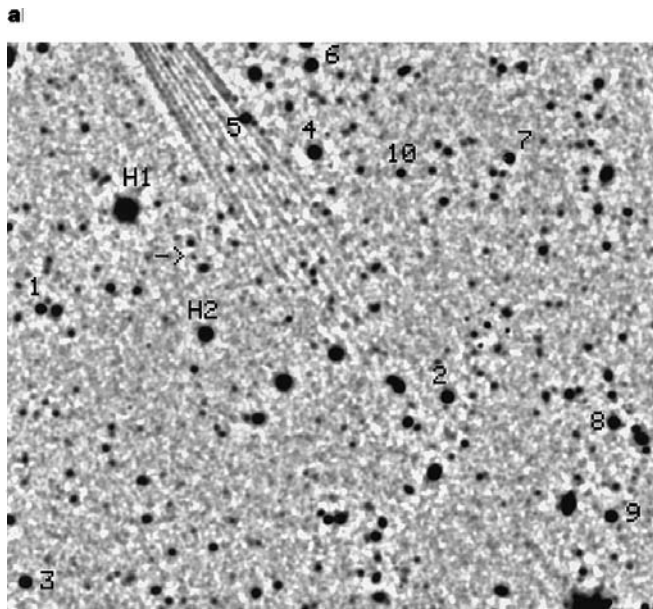
### Identifying the host galaxy

Using the above decay formula we predict  $R = 27.4 \pm 0.8$  mag on 1998 January 10 UT. In contrast, we find at the same epoch an object at the position of the OT with  $R = 25.6 \pm 0.17$  mag (see Fig. 1).

Furthermore, this object, which we will call ‘K’, is extended in comparison to stars in the same field. We suggest that K is the host galaxy of the GRB and has become apparent now that the OT has dimmed.

It is important to assess how well K coincides with the OT. To this end, we determined the location of the OT relative to 15 ‘tertiary’ stars in the vicinity of the OT in the December 16.52 I-band image, and likewise for K in the January 10 R-band image. The error in the measured angular difference between the OT and K is determined by two factors, as follows. (1) The error in determining the centroid of K, which primarily arises from the small number of signal photons and the sky background. We estimate this to be 0.20 pixel (r.m.s.); each LRIS pixel is a square of side 0.211 arcsec. The similar error for the OT is negligible because the OT is much brighter than K in the December 16 image that we used. (2) Errors due to coordinate transformation between the two images which are of different orientations. We determined a coordinate transformation between the frames using the measurements of the tertiary stars, and then transformed the centroid of the OT into the frame of January 10 image. This error turns out to be negligible compared to the error discussed in (1). The difference between the position of the OT in the transformed image and that of K is  $\Delta X = 0.24 \pm 0.20$  pixel =  $0.05 \pm 0.04$  arcsec and  $\Delta Y = 0.14 \pm 0.20$  pixel =  $0.03 \pm 0.04$  arcsec, corresponding to a radial separation of  $0.06 \pm 0.06$  arcsec.

One may argue that the density of galaxies as faint as K is



**Figure 1** Images of the field of the optical transient of GRB971214 and the associated host galaxy. **a**, I-band image of the field of the optical transient (OT) of GRB971214. The image is 3.5 arcmin by 3.0 arcmin and has been smoothed by a gaussian of full-width at half-maximum (FWHM) of 0.53 arcsec. The numbered stars are ‘secondary’ stars used for achieving consistency within our various data sets. Stars H1 and H2 allow us to link our photometry with that of Halpern *et al.*<sup>26</sup> and Henden *et al.*<sup>29</sup>. The ‘rays’ are light from the bright star (6.7 mag) SAO15663 diffracted by the telescope structure; this bright star is  $\sim 1$  arcmin northeast of the OT. The rays rotate as the telescope tracks the source. In **b**, the left panel is an I-band image of the OT obtained from data taken on 1997 December 16 UT. The image is a square of side 1 arcmin, and has been smoothed by a gaussian with FWHM of 0.53 arcsec. The OT is marked. The right panel is an R-band image of the field of OT obtained from data taken on 1998 January 10 UT. The image is a square of side 1 arcmin, and has been smoothed by a gaussian with FWHM of 0.53 arcsec. An extended object is seen at the position of the OT. We suggest that this is the host galaxy of GRB971214. For all three images, North is to the top and East to the left. In all three images the optical transient is marked by an arrow. We have also obtained the spectrum of the galaxy  $\sim 4.5$  arcsec to the northeast of the OT. The spectrum shows a prominent [O II] 3,727 emission line and the usual absorption features at a redshift of 0.5023.



sufficiently high that the coincidence of the OT with a background galaxy is not improbable. The cumulative surface density<sup>35</sup> of galaxies with  $R \leq 25.5$  mag is  $3.9 \times 10^5$  per degree<sup>2</sup>. Considering angular offsets up to the apparent radius of the galaxy, 0.35 arcsec (see Fig. 1), the chance coincidence probability is reasonably small ( $\sim 10^{-5}$ ). We therefore suggest that galaxy K is the host of the OT and thus is the host of GRB971214.

### Determining the redshift

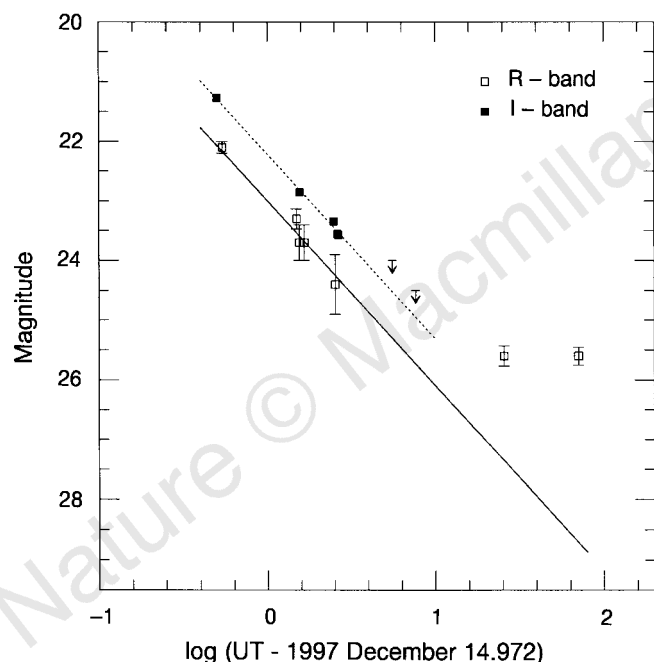
Spectroscopic observations of the OT and its host galaxy were obtained using LRIS<sup>24</sup>. Our initial attempt to obtain the spectrum of the OT on 1997 December 17 UT was unsuccessful, largely due to the bright moonlight. Subsequent observations were obtained under much better sky conditions, and after the OT had faded and object K dominated the light. Almost all spectra were obtained with the slit position angle close to the parallactic angle, and the wavelength-dependent slit losses are not important for the discussion below. The log and the technical details of the spectroscopic observations can be found in Table 2.

The spectrum, shown in Fig. 3, shows a prominent emission line

at 5,382.1 Å with a clear continuum drop immediately on the blue (short-wavelength) side of the line. No other emission line is detected. The continuum rises slightly towards the red and is undetectable at about 4,000 Å.

There are only two plausible interpretations of this prominent emission line. All other choices would require a presence of other, stronger emission lines which are not seen in our spectrum. First, the line could be [O II] 3,727 at  $z = 0.444$ . We would then expect to see comparably strong [O III] 5,007 and 4,959 lines at 7,230 Å and 7,160 Å, H $\beta$  4,861 Å at 7,020 Å, and perhaps also other Balmer lines. Yet none of these lines are seen in our data, even though they fall in a reasonably clean part of the spectrum. It would be also hard to understand the abrupt drop in the continuum blueward of the emission line.

The alternative interpretation, and the interpretation which we favour, is that the emission line is Ly $\alpha$  1,215.7 Å at a redshift  $z_{\text{em}} = 3.428$ . The overall appearance of the spectrum is typical of the known, star-forming galaxies at comparable redshifts<sup>19–21</sup>. In the absence of an active nucleus, no other strong emission lines would be expected within the wavelength range probed by our spectrum.



**Figure 2** The R- and I-band light curve of the OT of GRB971214. The OT is identified in Fig. 1. The x-axis is  $\log(t - t_0)$  where  $t$  is the UT date of the observation and  $t_0$  is the UT date of the GRB. All times are measured in units of days. The GRB took place on 1997 December 14.9727 (ref. 16). At I band the OT is well detected for the first three nights. Upper limits exist for the nights of UT 1997 December 20 and December 22. The dotted line is a linear least-squares fit to the I-band data but restricted to these first three nights. The assumed model is  $I(t - t_0) = I_1 + s \log(t - t_0)$ ; here  $I$  is the magnitude at time  $t$ ,  $s$  is the slope and  $I_1$  the offset. The fits yield  $I_1 = 22.22 \pm 0.18$  and  $s = 3.09 \pm 0.52$ . For reasons discussed in the text it is reasonable to assume that the R-band light curve has the same slope,  $s$ . The solid line is parallel to the dotted line but goes through our R-band magnitude of 1997 December 16.63 UT. This line describes adequately the R-band light curve for the first three nights. Clearly, the December 15.50 point of Henden *et al.*<sup>29</sup> is not well accounted for by this model. The R-band measurements of 1998 January 10 and February 24 lie well above the extrapolation of the dotted line. We attribute this excess over the decaying optical transient as arising from galaxy K shown in Fig. 1. In both the January 10 and February 24 images, the full width at half maximum (FWHM) of K is 1.07 arcsec which is larger than the same estimated from stars in the vicinity of K (Table 1). The rough size of K is thus  $\sim 0.65$  arcsec.

**Table 1** Summary of R- and I-band measurements

| Epoch (UT)      | Band | Magnitude        | Notes and refs*                               |
|-----------------|------|------------------|-----------------------------------------------|
| 1997 Dec. 15.47 | I    | $21.27 \pm 0.30$ | Ref. 26                                       |
| 1997 Dec. 16.52 | I    | $22.85 \pm 0.33$ | $9 \times 120$ s, 0.9", R.G.                  |
| 1997 Dec. 17.45 | I    | $23.34 \pm 0.31$ | $9 \times 120$ s, 1.1", R.G.†                 |
| 1997 Dec. 17.60 | I    | $23.55 \pm 0.26$ | $11 \times 120$ s, 0.8", R.G.                 |
| 1997 Dec. 17.63 | I    | $23.57 \pm 0.40$ | $11 \times 120$ s, 0.63", R.G.‡               |
| 1997 Dec. 20.53 | I    | >24              | $10 \times 360$ s, A.V.F. 1.4"§               |
| 1997 Dec. 22.63 | I    | >24.5            | $2 \times 400$ s, 0.9", M.D.¶                 |
| 1997 Dec. 15.50 | R    | $21.7 \pm 0.10$  | Ref. 29                                       |
| 1997 Dec. 15.51 | R    | $22.1 \pm 0.10$  | Ref. 28                                       |
| 1997 Dec. 16.46 | R    | $23.3 \pm 0.17$  | $6 \times 180$ s, 0.86", R.G.                 |
| 1997 Dec. 16.63 | R    | $23.7 \pm 0.10$  | $3 \times 180$ s, 0.56", R.G.‡                |
| 1997 Dec. 16.62 | R    | $23.7 \pm 0.30$  | Ref. 28                                       |
| 1997 Dec. 17.51 | R    | $24.4 \pm 0.50$  | Ref. 25                                       |
| 1998 Jan. 10.62 | R    | $25.6 \pm 0.17$  | $12 \times 300$ s, 0.86", J. Aycock¶          |
| 1998 Feb. 24.52 | R    | $25.6 \pm 0.15$  | $10 \times 300$ s, 0.87", S.R.K.#             |
| 1998 Feb. 24.58 | B    | >26.8            | $6 \times 300, 3 \times 600, 1.00"$ , S.R.K.# |

\* The entries in the last column are: (for Keck observations) the number of frames obtained, the integration time, the seeing specified as full-width at half-maximum (FWHM), the observer or the reference (for observations reported in the literature). Additional notes are indicated by a letter. The log of the December 15 'guide' camera observations is not included here as we did not use those data in any quantitative analysis. (The OT was detected in those data, though.) Successive frames are displaced by 5–15 arcsec and the final image is obtained by registering the frames and then adding. All the imaging analysis was carried out with IRAF, a software package supplied by the National Optical Astronomy Observatories. † Rhoads<sup>27</sup> reports  $I = 22.9 \pm 0.4$  on 1997 Dec 17.37 UT. This is consistent, within errors, with our measurement of the same night. However, due to the large error bars, this is not included in our compilation (above).

‡ This observation was affected by light diffracted by the telescope structure contaminating the region in the vicinity of the OT (see Fig. 1). The diffracted pattern was  $\sim 15$  pixels wide and long. An image was formed by assigning each pixel a value equal to the median of a box  $75 \times 5$  pixels, aligned along the diffraction spikes. All features smaller than this box disappear in the resultant image, leaving only the diffracted rays. This median image was then subtracted from the original image leaving an image free from the diffracted pattern. Photometry of the OT, reported in this table, was then performed on this subtracted image. § The data were taken in the polarimetric mode of LRIS and hence the light is split into two orthogonal polarization components—we call this the top and bottom channels. The images from each of the channels were reduced separately, averaged together and the two resultant images were co-added to give the final image. The seeing as estimated by observations of the standard star SA 104 440 (ref. 30) was exceptionally bad,  $\sim 1.4$  arcsec. No object was found in the final image at the position of the OT. We do not have a good estimate of the true point-spread function in the final image as the useful field is only 25 arcsec across and there are no bright stars within this. Based on noise statistics within an aperture of radius 1.5 arcsec we set a  $3\sigma$  detection limit of  $I = 24$  mag for this data. Analysis of artificial stars with FWHM of 1.4 arcsec using the IRAF task DAOFIND led to a detection limit of 23 mag. However, visual inspection of the artificial stars knowing the position of the object results in a more stringent limit, closer to the  $I = 24$  mag obtained above.

¶  $3\sigma$  upper limit estimated from the detection statistics of artificial stars which have the same point-spread function as that of nearby stars and embedded in the general vicinity of the OT. This observation suffered from high sky brightness (19.1 mag per square arcsec) and also lack of numerous exposures.

# The first four exposures suffered from brighter sky as the Moon was still up in the sky. The image shown in Fig. 1 is the sum of the next eight exposures.

#  $3\sigma$  upper limit. See footnote ¶ for details of the procedure of deriving the upper limit.

Additional arguments in favour of the large-redshift interpretation include:

(1) The drop across the Ly $\alpha$  line, due to the absorption by intervening hydrogen, is characteristic of objects at such high redshifts. The measured drop amplitude<sup>36</sup>  $D_A = 0.35 \pm 0.05$  is exactly the mean value seen in the spectra of quasars at this redshift<sup>37</sup>.

(2) We expect no flux blueward of 4,030 Å, the redshifted Lyman continuum break. Additionally, we expect a relatively flat, star-formation-powered continuum redward of the Ly $\alpha$  line. Our spectrum is consistent with these expectations. The expected suppression of blue light is also supported by our imaging observations in the B band (Table 1); this band covers the range 3,900–4,800 Å. Our  $3\sigma$  upper limit on the B magnitude of galaxy K is 26.8 mag, corresponding to a B-band flux of  $<84$  nJy. We find that the red portion of the spectrum can be approximated by  $F_\nu = 174(\nu/\nu_R)^\alpha$  nJy with  $\alpha = -0.7 \pm 0.2$ ; here  $F_\nu$  is the spectral density at frequency  $\nu$  and  $\nu_R = 4.7 \times 10^{14}$  Hz, the centre frequency of the R band. The extrapolation of this spectrum to the B band ( $\nu_B = 6.8 \times 10^{14}$  Hz) is 134 nJy, above our upper limit.

(3) We plot in Fig. 3 the expected locations of some of the commonly observed interstellar gas absorption lines seen in the spectra of star-forming galaxies at high redshifts<sup>19–21</sup>. We note several coincidences with dips in the observed spectrum.

If this interpretation is correct, a better estimate of the galaxy's redshift can be obtained as follows. The Ly $\alpha$  emission line tends to be shifted systematically to the red in such galaxies, due to resonant scattering and absorption in the ambient gas. In any given case, the bias thus introduced in the centring of the peak of the Ly $\alpha$  emission

depends on the exact geometry and kinematics of the gas and dust, which are not known; but on average the effect is to shift the peak of the Ly $\alpha$  emission to the red, relative to the systemic redshift of the galaxy<sup>19–21</sup>. We measure the wavelength of the apparent absorption dip immediately on the high-frequency side of the emission line as:  $\lambda_{\text{obs,air}} = 5,356.1$  Å. Taking the mean of this wavelength and that of the emission line, and applying the standard air-to-vacuum correction, we obtain for the systemic redshift of this galaxy,  $z_K = 3.418 \pm 0.010$ .

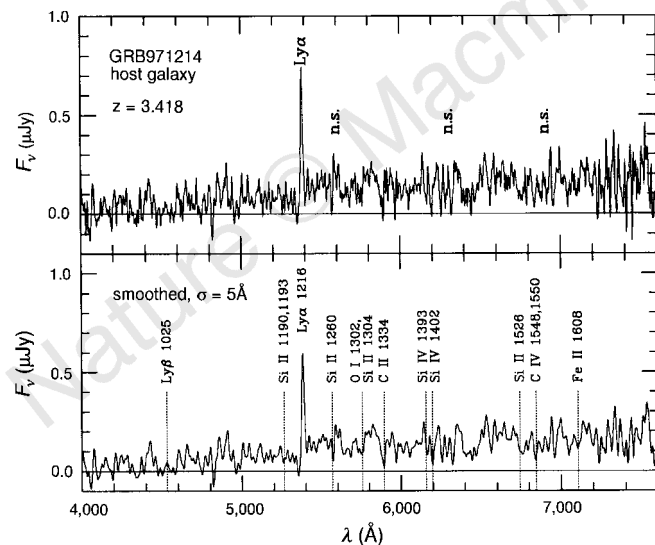
As discussed above the continuum slope of the observed spectrum redward of the emission line is  $F_\nu \propto \nu^\alpha$  with  $\alpha = -0.7 \pm 0.2$ . Unobscured star-forming galaxies have  $\alpha \approx 0.0$  to  $-0.5$ , depending on the initial mass function and the history of star formation. The slightly steeper slope of our spectrum suggests a modest amount of rest-frame extinction for the galaxy as a whole. This of course does not constrain the extinction along any particular line of sight, such as the direction to the OT itself. We note that similar extinctions are inferred for star-forming galaxies at comparable redshift<sup>19</sup>.

Two corrections must be taken into account before a quantitative interpretation of the spectrum can be done. First, the Galactic reddening in this direction is estimated<sup>33</sup> to be  $E_{B-V} \approx 0.016$  mag, implying a Galactic extinction correction to the observed fluxes of about 6% at the wavelength of the emission line, and about 4% in the R band. Second, by comparing the spectrum to our measured R-band photometry of the galaxy, we estimate the loss due to the finite size of the slit to be a factor of 1.48.

### The nature of the host galaxy

We now discuss the physical properties and nature of this galaxy. Assuming a standard Friedman model cosmology with  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $\Omega_0 = 0.3$ , we derive a luminosity distance ( $d_L$ ) of  $9.7 \times 10^{28} \text{ cm}$  (changing the cosmological parameters to  $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $\Omega_0 = 1$  yields  $d_L = 8.6 \times 10^{28} \text{ cm}$ ). ( $H_0$ , the Hubble constant, is the expansion rate of the Universe at the present time.  $\Omega_0$  is the ratio of the mean density of the Universe to the closure density.) Assuming that the intrinsic spectrum  $F_\nu$  is  $\propto \nu^{-0.7}$  and scaling from the observed R-band magnitude corrected for the Galactic extinction, we find the rest-frame B-band flux to be  $\sim 0.45 \mu\text{Jy}$ . The corresponding absolute B magnitude is  $M_B \approx -20.9$ , and is about equal to the absolute magnitude of a typical,  $L_*$  galaxy today. Given its high redshift, this galaxy probably has yet to produce most of its stars, and even with some evolutionary fading of its present-day luminosity it may evolve to an  $L_*$  galaxy.

The observed Ly $\alpha$  emission line flux is  $F_{\text{Ly}\alpha} = (3.8 \pm 0.4) \times 10^{-18} \text{ erg cm}^{-2} \text{ s}^{-1}$ . Correcting for the flux calibration zero-point



**Figure 3** The composite Keck spectrum of the host galaxy of GRB971214. The top panel shows the original data, with the locations of the prominent night sky (n.s.) emission lines indicated. The bottom panel shows the same spectrum smoothed with a Gaussian with  $\sigma = 5$  Å, which is approximately equal to the effective instrumental resolution. Locations of several absorption features commonly seen in the spectra of  $z \approx 3$  galaxies are indicated. The redshift,  $z = 3.418$ , has been derived from the mean point of the Ly $\alpha$  emission and absorption features, as described in the text. The log of the observations and the details of the spectrograph settings can be found in Table 2. The 'lower' resolution data were reduced completely independently by two of the authors (S.G.D. and K.L.A.), using independent reduction packages. The results are in an excellent mutual agreement. The 'higher' resolution data are also fully consistent with them, showing essentially the same spectroscopic features. The useful wavelength range spanned by the lower-resolution data is  $\sim 4,000$ – $7,600$  Å, and the higher-resolution data spans  $\sim 4,900$ – $7,300$  Å. All of the spectra have been averaged with appropriate signal-to-noise weighting, after suitable resampling.

**Table 2 Summary of spectroscopic observations**

| Epoch (UT)   | Grating (lines per mm) | Slit width (arcsec) | Int. time (s)     | Notes  |
|--------------|------------------------|---------------------|-------------------|--------|
| 1997 Dec. 28 | 300                    | 1.5                 | $2 \times 1800$ s | T.K.   |
| 1998 Feb. 3  | 300                    | 1.0                 | $5 \times 1800$ s | S.R.K. |
| 1998 Feb. 22 | 600                    | 1.0                 | $2 \times 1800$ s | S.R.K. |
| 1998 Feb. 23 | 600                    | 1.0                 | $6 \times 1800$ s | S.R.K. |

Entries (from left to right) are: the UT date of the observation, the grating used, the slit width in arcsec, the number of exposures and the integration time per exposure, and the name of the principal observer. The 300 lines per mm grating is blazed to 5,000 Å and the blaze wavelength of the 600 lines per mm grating is 7,500 Å. The centre wavelength of the 300 lines per mm spectra (hereafter 'lower' resolution data) is  $\sim 6,400$  Å and the dispersion is  $\sim 2.45$  Å per pixel. The centre wavelength of the 600 lines per mm spectra, the 'higher resolution' data, is 6,100 Å and the dispersion is 1.25 Å per pixel. In both cases, the spectral resolution as parametrized by the FWHM is  $\sim 5$  pixels. The spectrum of the standard star HZ 44 (ref. 45) obtained on February 3 was used to flux calibrate the lower-resolution spectra, and the spectrum of GD 153 (ref. 46), observed on February 22, was used to calibrate the higher-resolution spectra. The zero-point uncertainty of the flux calibration is estimated to be  $\sim 10\%$ , judging by the internal agreement, but the slit losses are likely to be higher. Wavelength calibration was derived from arc lamp spectra taken immediately after the observations of the target. The random errors in the wavelength calibration are 0.3 Å and the estimated systematic errors due to instrument flexure are of the same order. Both these uncertainties are unimportant for our redshift determination discussed in the text.

and the Galactic extinction, this becomes  $F_{\text{Ly}\alpha} = (6.2 \pm 0.7) \times 10^{-18} \text{ erg cm}^{-2} \text{ s}^{-1}$ . For our assumed cosmology, the implied Ly $\alpha$  line luminosity is  $L_{\text{Ly}\alpha} = (7.3 \pm 0.8) \times 10^{41} \text{ erg s}^{-1}$ . Estimates of conversion of the Ly $\alpha$  line luminosity to the implied unobscured star-formation rate (SFR) are in the range  $L_{\text{Ly}\alpha} = (7 \pm 4) \times 10^{41} \text{ erg s}^{-1}$  for  $\text{SFR} = 1 M_{\odot} \text{ yr}^{-1}$  (where  $M_{\odot}$  is the solar mass), depending on the stellar initial mass function (IMF)<sup>39,40</sup>. We thus estimate the unobscured star-formation rate in this galaxy to be approximately  $(1.0 \pm 0.5) M_{\odot} \text{ yr}^{-1}$ .

However, the Ly $\alpha$  emission line is probably attenuated by resonant scattering in neutral hydrogen and by absorption by dust. An independent, and perhaps more robust, estimate of the star-formation rate can be obtained from the rest-frame continuum luminosity at 1,500 Å (ref. 41). We note that the observed flux at a wavelength of  $1,500(1 + z_K)$  Å is 0.22 μJy. This translates to a star-formation rate, under the usual assumption of a Salpeter IMF, of  $5.2 M_{\odot} \text{ yr}^{-1}$ . These numbers should be regarded as lower limits given the unknown extinction in the rest-frame of galaxy K.

On the whole, the properties of galaxy K are typical of the known systems at comparable redshifts<sup>19–21</sup>, thus giving us some confidence that our redshift interpretation is indeed correct.

The fluence of this GRB<sup>17</sup> above the observed photon energy of 20 keV is  $F = 1.1 \times 10^{-5} \text{ erg cm}^{-2}$ . The  $\gamma$ -ray fluence as observed by the Gamma-ray Burst Monitor (GRBM) on board BeppoSAX is  $(0.9 \pm 0.9) \times 10^{-5} \text{ erg cm}^{-2} \text{ s}^{-1}$  (above 40 keV). The burst was also observed by the All Sky Monitor on board the X-ray satellite XTE (ref. 42). D. A. Smith (personal communication) estimates the fluence in the 2–12 keV band to be  $1.8(\pm 0.03) \times 10^{-7} \text{ erg cm}^{-2}$ . Thus the isotropic energy loss in  $\gamma$ -rays alone at the distance to the host is  $4\pi d_L^2 F/(1+z) \approx 3 \times 10^{53} \text{ erg}$ .

The currently favoured model<sup>43</sup> for GRBs is coalescence of neutron stars. The coalescence is expected to release most of the energy in neutrinos (as in type II supernovae) and about  $10^{51} \text{ erg}$  is supposed to be in the form of electromagnetic energy. The measured fluence of GRB971214 when combined with our proposed redshift for the GRB appears to be inconsistent with the expectations of the neutron-star merger model in its simplest form; however, it is possible that more elaborate versions of this model could be made to fit these observations. Non-spherical emission could reduce the strain on the energy budget but will not alter the fact that this fairly bright burst originated from such a large redshift.

The most significant implication of our hypothesis that K is the host galaxy of GRB971214 is the implied extreme energetics. The energy budget for GRBs goes up from the traditional  $10^{51} \text{ erg}$  to perhaps the entire energy available in the coalescence of neutron-star mergers<sup>43</sup>. Other energetic models, ranging from the death of an extremely massive star<sup>23</sup> to coalescence of black-hole neutron star binaries<sup>22</sup> may then become more attractive. In the latter models, GRBs are directly related to the formation rate of massive stars. If that is the case, then the typical redshift of GRBs is approximately 2 (ref. 44), consistent with the suggested high redshift for GRB971214. Moreover, this burst evidently did not originate in a highly obscured star-forming region in its host galaxy.

Regardless of the details of their genesis, GRBs appear to be the brightest known objects in the Universe, albeit over the limited duration of the burst. The term “hypernova”<sup>23</sup> can be justifiably used to describe these most extreme events, especially their afterglow. GRB971214 was not a particularly faint event, and thus statistically we expect many fainter events to arise from larger redshifts. The high brightness of GRBs and their optical transients offer us exciting and new opportunities to probe the Universe at high redshifts. □

Received 18 March; accepted 14 April 1998.

- Klebesadel, R. W., Strong, I. B. & Olson, R. A. Observations of gamma-ray bursts of cosmic origin. *Astrophys. J.* **182**, L85–L88 (1973).
- Boella, G. *et al.* BeppoSAX, the wide band mission for x-ray astronomy. *Astron. Astrophys. Suppl. Ser.* **122**, 299–399 (1997).
- Costa, E. *et al.* Discovery of an X-ray afterglow associated with the  $\gamma$ -ray burst of 28 February 1997. *Nature* **387**, 783–785 (1997).
- van Paradijs, J. *et al.* Transient optical emission from the error box of the  $\gamma$ -ray burst of 28 February 1997. *Nature* **386**, 686–689 (1997).
- Frail, D. A., Kulkarni, S. R., Nicastro, L., Feroci, M. & Taylor, G. B. The radio afterglow from the  $\gamma$ -ray burst of 8 May 1997. *Nature* **389**, 261–263 (1997).
- Metzger, M. R. *et al.* Spectral constraints on the redshift of the optical counterpart to the  $\gamma$ -ray burst of 8 May 1997. *Nature* **387**, 878–880 (1997).
- Bond, H. E. *et al.* *IAU Circ.* No. 6665 (1997).
- Djorgovski, S. G. *et al.* The optical counterpart to the  $\gamma$ -ray burst 970508. *Nature* **387**, 876–878 (1997).
- Costa, E. *et al.* *IAU Circ.* No. 6649 (1997).
- Piro, L. *et al.* *IAU Circ.* No. 6656 (1997).
- Bloom, J. S., Kulkarni, S. R., Djorgovski, S. G. & Frail, D. A. *GCN Note* No. 30 (1998).
- Zharikov, S. V., Sokolov, V. V. & Baryshev, Y. V. *GCN Note* No. 31 (1998).
- Galama, T. J. *et al.* Optical followup of GRB 970508. *Astrophys. J.* (submitted).
- Sahu, K. C. *et al.* Observations of GRB 970228 and GRB 970508 and the neutron star merger mode. *Astrophys. J.* **489**, L127–L131 (1997).
- Halpern, J. P., Thorstensen, J. R., Helfand, D. J. & Costa, E. Optical afterglow of the  $\gamma$ -ray burst of 14 December 1997. *Nature* **393**, 41–43 (1998).
- Heise, J. *et al.* *IAU Circ.* No. 6787 (1997).
- Kippen, R. M. *et al.* *IAU Circ.* No. 6789 (1997).
- Antonelli, L. A. *et al.* *IAU Circ.* No. 6792 (1997).
- Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M. & Adelberger, K. Spectroscopic confirmation of a population of normal star-forming galaxies at redshifts  $z > 3$ . *Astrophys. J.* **462**, L17–L21 (1996).
- Steidel, C. C., Giavalisco, M., Dickinson, M. & Adelberger, K. L. Spectroscopy of Lyman break galaxies in the Hubble Deep Field. *Astron. J.* **112**, 352–358 (1996).
- Steidel, C. C. *et al.* A large structure of galaxies at redshift  $z \sim 3$  and its cosmological implications. *Astrophys. J.* **492**, 428–438 (1998).
- Mezáros, P. & Rees, M. J. Poynting jets from black holes and cosmological gamma-ray bursts. *Astrophys. J.* **482**, L29–L31 (1997).
- Paczynski, B. Are gamma-ray bursts in star-forming regions? *Astrophys. J.* **492**, L45–L48 (1998).
- Oke, J. B. *et al.* The Keck low-resolution imaging spectrometer. *Publ. Astron. Soc. Pacif.* **107**, 375–385 (1995).
- Castander, F. J. *et al.* *GCN Note* No. 11 (1997).
- Halpern, J., Thorstensen, J., Helfand, D. & Costa, E. *IAU Circ.* No. 6788 (1997).
- Rhoads, J. *IAU Circ.* No. 6793 (1997).
- Diercks, A. *et al.* *IAU Circ.* No. 67921 (1997).
- Henden, A. A., Luginbuhl, C. B. & Vrba, F. J. *GCN Note* No. 16 (1997).
- Landolt, A. U. UBVR photometric standard stars in the magnitude range  $11.5 < V < 16.0$  around the celestial equator. *Astron. J.* **104**, 340–371 (1992).
- Wijers, R. A. M. J., Rees, M. J. & Mezáros, P. Shocked by GRB 970228: the afterglow of a cosmological fireball. *Mon. Not. R. Astron. Soc.* **288**, L51–L56 (1997).
- Sokolov, V. V. *et al.* BVR<sub>Ic</sub> photometry of GRB 970508 optical remnant: May–August, 1997. *Astron. Astrophys.* (in the press); preprint <http://xxx.lanl.gov, astro-ph/0902341> (1998).
- Galama, T. J. *et al.* Optical follow-up of GRB 970508. *Astrophys. J.* (in the press); preprint <http://xxx.lanl.gov, astro-ph/9802160> (1998).
- Waxman, E. Gamma-ray-burst afterglow: supporting the cosmological fireball model, constraining parameters, and making predictions. *Astrophys. J.* **485**, L5–L8 (1997).
- Hogg, D. W. *et al.* Counts and colors of faint galaxies in the U and R bands. *Mon. Not. R. Astron. Soc.* **288**, 404–410 (1997).
- Oke, J. B. & Korycansky, D. Absolute spectrophotometry of very large redshift quasars. *Astrophys. J.* **255**, 11–19 (1996).
- Kennefick, J. D., Djorgovski, S. G. & de Carvalho, R. R. The luminosity function of  $z > 4$  quasars from the second Palomar sky survey. *Astron. J.* **110**, 2553–2565 (1995).
- Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust IR emission for use in estimation of reddening and CMBR foregrounds. *Astrophys. J.* (in the press); preprint <http://xxx.lanl.gov, astro-ph/0910327>.
- Thompson, D., Djorgovski, S. & Trauger, J. A narrow-band imaging survey for primeval galaxies. *Astron. J.* **110**, 963–981 (1995).
- Charlot, S. & Fall, S. M. Lyman-alpha emission from galaxies. *Astrophys. J.* **415**, 580–588 (1993).
- Leitherer, C., Robert, C. & Heckman, T. M. Atlas of synthetic ultraviolet-spectra of massive star populations. *Astrophys. J. Suppl.* **99**, 173–187 (1995).
- Doty, J. P. The All Sky Monitor for the X-ray Timing Explorer. *Proc. SPIE* **982**, 164–172 (1988).
- Narayan, R., Paczyński, B. & Piran, T. Gamma-ray bursts as the death throes of massive binary stars. *Astrophys. J.* **395**, L83–L86 (1992).
- Wijers, R. A. M. J., Bloom, J., Bagla, J. S. & Natarajan, P. Gamma-ray bursts from stellar remnants: probing the universe at high redshift. *Mon. Not. R. Astron. Soc.* **294**, L13–L17 (1998).
- Massey, P., Strobel, K., Barnes, J. & Anderson, E. Spectrophotometric standards. *Astrophys. J.* **328**, 315–333 (1988).
- Bohlin, R., Colina, L. & Finley, D. White dwarf standard stars: G191-B2B, GD 71, GD 153, HZ 43. *Astron. J.* **110**, 1316–1325 (1995).

**Acknowledgements.** The observations reported here were obtained at the W. M. Keck Observatory, which is operated by the California Association for Research in Astronomy, a scientific partnership among California Institute of Technology, the University of California and NASA. It was made possible by the financial support from W. M. Keck Foundation. We thank W. Sargent, Director of the Palomar Observatory, F. Chaffee, Director of the Keck Observatory and our colleagues for continued support of our GRB program. We thank J. C. Clemens and M. H. van Kerkwijk for help with observations and exchange of dark time. S.R.K.’s research is supported by the NSF and NASA. S.G.D. acknowledges partial support from the Bressler Foundation. A.N.R. is grateful to the International Astronomical Union for a travel grant.

Correspondence and requests for materials should be addressed to S.R.K. (e-mail: srk:surya.caltech.edu).



# Optical afterglow of the $\gamma$ -ray burst of 14 December 1997

J. P. Halpern\*, J. R. Thorstensen†, D. J. Helfand\* & E. Costa‡

\* Columbia University, 550 West 120th Street, New York, New York 10027, USA

† Dartmouth College, 6127 Wilder Laboratory, Hanover, New Hampshire 03755, USA

‡ Istituto di Astrofisica Spaziale CNR, 00044 Frascati, Italy

The very recent detection of the faint host galaxy of one  $\gamma$ -ray burst<sup>1–4</sup>, and the determination of a cosmological redshift for another<sup>5</sup>, demonstrates that these events are the most luminous phenomena in the Universe, emitting more energy in radiation than a supernova over just a few seconds. The source of this energy is still unknown, but may become clear through studies of the counterparts at longer wavelengths. Here we report the detection of an optical counterpart to a  $\gamma$ -ray burst (GRB971214) that occurred on 14 December 1997. It faded rapidly over a two-week period, just like the previous two optical transients<sup>1,6–11</sup>, which dispels any doubt that the three events are the optical afterglows of  $\gamma$ -ray bursts. The 14 December optical transient is the faintest of the three, and also is much redder than the other two. This reddening probably arises because of scattering by interstellar dust along the line of sight, which is presumably present in the denser regions of the host galaxy, where stars form. This suggests that the burst's progenitor did not stray too far from the point of its birth, which, regardless of the nature of the source, appears to be in a region of dense gas.

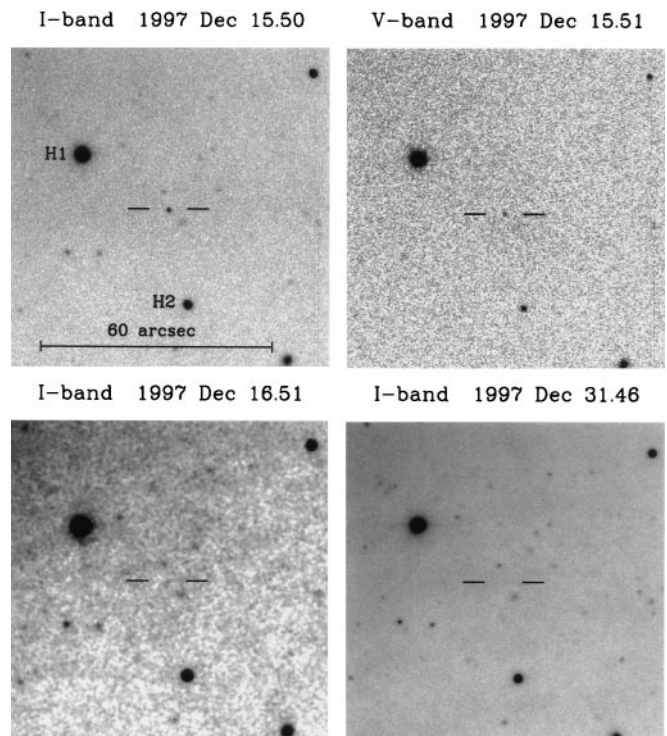
A 3.9-arcmin-radius error circle for GRB971214, which occurred on 1997 December 14.9727 UT, was obtained by the Wide Field Camera on the BeppoSAX X-ray satellite<sup>12</sup>. The first optical images of the field were acquired 12 h after the event by one of us (J.R.T.) observing on the 2.4-m Hiltner telescope of the MDM Observatory on Kitt Peak. A series of CCD images were recorded in the V and I bands beginning on 1997 December 15.47 UT. On the following night, repeat I-band observations resulted in the recognition of a fading transient within the error circle, and preliminary magnitudes were reported<sup>13</sup>. Table 1 gives our final MDM magnitudes based on detailed analysis of all our data. Preliminary magnitudes reported from other telescopes<sup>14–17</sup> are also listed. All the MDM observations were photometric, and were calibrated with the same set of standard stars<sup>18</sup>. The 15 and 16 December data were near full Moon, and I-band magnitudes of  $21.30 \pm 0.04$  and  $23.0 \pm 0.4$ , respectively, were measured from the nightly averages. Deeper I-band images were obtained at MDM on 31 December, a moonless night, on which an upper limit of  $I > 24.7$  mag was measured. Thus, a fading by at least 3.4 mag was observed over 17 d. Figure 1 shows the nightly sums of the MDM images.

All three of the GRB optical afterglows detected so far were accompanied by X-ray afterglows<sup>19–21</sup>. In the case of GRB971214, the position of the optical transient (see Fig. 1) lies 71 arcsec from the position of the X-ray afterglow in the SAX narrow-field

instrument<sup>20</sup>, which was reported with an uncertainty of  $\sim 1$  arcmin. As SAX position uncertainties are now dominated by systematic effects of attitude reconstruction using the satellite's new 1-gyro mode, we regard the fading X-ray source 1SAX J1156.4+6513 as coincident with the optical transient.

Figure 2 shows the light curves of GRB971214 assembled from the measurements listed in Table 1. The  $R$ ,  $I$  and  $J$  data are broadly consistent with the same rate of decay, a power law with index  $-1.4 \pm 0.2$ , although the 17 December I-band point from the KPNO 0.9-m telescope may be somewhat discrepant. This decay rate is roughly consistent with those of the previous optical afterglows, GRB970228 and GRB970508, for which indices of  $-1.1$  and  $-1.2$ , respectively, were derived<sup>10,11,22,23</sup>. Our 15 December observations can also be used to set an upper limit on rapid variability. Although we obtained nine I-band images over a 1.5-h period, the dispersion among the measurements was  $< 0.1$  mag, which we consider to be a conservative limit on intrinsic variability on this timescale. The quoted error of 0.04 on the 15 December I-band magnitude is the error in the mean of these measurements.

GRB971214 is both fainter and redder than the previous two optical afterglows. Comparing the three events at an observed time of 2 d, when each had definitely entered its decay phase, the interpolated I-band magnitudes were 19.3,  $\sim 22.1$  and 23.4, respectively, for GRB970508<sup>9–11,22</sup>, GRB970228<sup>23,24</sup> and GRB971214. The colour  $V - I = 1.35 \pm 0.12$  that we measure for GRB971214 at a time of 12 h contrasts with the other GRBs, which were observed initially to have  $V - I \approx 0.7$ – $0.9$ . In Fig. 3, we convert the VRI magnitudes obtained around December 15.5 to fluxes<sup>25</sup>. Galactic



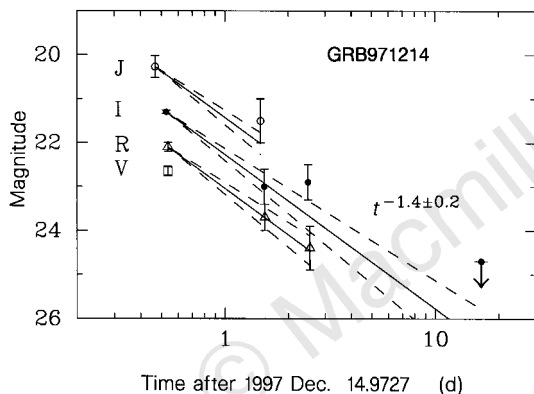
**Figure 1** MDM Observatory images of the optical counterpart of GRB971214. All were obtained using a SiTe 2048  $\times$  2048 CCD, with 0.27-arcsec pixels. North is up, and east is to the left. The December 16.51 image (only) was convolved with a two-dimensional gaussian with  $\sigma = 1$  pixel. The position of the transient is right ascension 11 h 56 min 26.40 s declination  $65^\circ 12' 00.5''$  (J2000), measured with respect to the USNO A1.0 reference system<sup>36</sup>. The estimated uncertainty is 0.75 arcsec in radius. Labelled comparison stars are H1 ( $I = 15.97$  mag,  $V - I = 0.87$ ) and H2 ( $I = 18.55$  mag,  $V - I = 2.66$ ). Many of the fainter objects are galaxies, including two that are within 5 arcsec of the transient. The compact one to the NNE has  $I = 22.8 \pm 0.1$  mag, and the more extended one to the southwest has  $I = 22.0 \pm 0.2$  mag.

**Table 1** Photometry of GRB971214

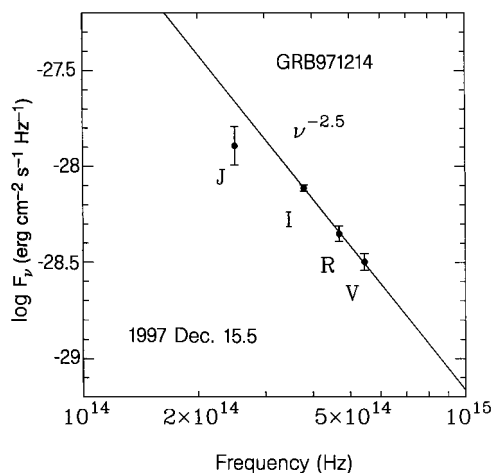
| Date (1997 UT) | Magnitude            | Telescope  | Ref. |
|----------------|----------------------|------------|------|
| Dec. 15.51     | $V = 22.65 \pm 0.11$ | MDM 2.4-m  |      |
| Dec. 15.51     | $R = 22.1 \pm 0.01$  | ARC 3.5-m  | 14   |
| Dec. 16.52     | $R = 23.7 \pm 0.3$   | ARC 3.5-m  | 14   |
| Dec. 17.51     | $R = 24.4 \pm 0.5$   | ARC 3.5-m  | 15   |
| Dec. 15.50     | $I = 21.30 \pm 0.04$ | MDM 2.4-m  |      |
| Dec. 16.51     | $I = 23.0 \pm 0.4$   | MDM 2.4-m  |      |
| Dec. 17.46     | $I = 22.9 \pm 0.4$   | KPNO 0.9-m | 16   |
| Dec. 31.46     | $I > 24.7$           | MDM 2.4-m  |      |
| Dec. 15.44     | $J = 20.27 \pm 0.25$ | ARC 3.5-m  | 17   |
| Dec. 16.45     | $J \approx 21.5$     | ARC 3.5-m  | 17   |

extinction in this direction is negligible. It can be seen that the *VRI* points for GRB971214 fall along a power-law spectrum of slope  $\alpha = -2.5 \pm 0.2$ , with some downward curvature in the J band. Within the larger errors of the 16 December data, there is no evidence for a change in the shape of the optical spectrum. In contrast, the spectral slope of GRB970508 was found to be  $\alpha = -1.1$  in the photometry<sup>22</sup>, and  $-0.9 \pm 0.3$  in the spectroscopy<sup>5</sup>. If interpreted as a synchrotron spectrum in the fireball model, the steep optical spectrum of GRB971214 is a probable indicator of reddening, which could be attributable to the environment of the host galaxy in the immediate neighbourhood of the transient.

Evidence that the steep optical spectrum is due to reddening, as opposed to synchrotron losses by the emitting electrons, comes from the observation of an X-ray afterglow<sup>20</sup> in the first day after the burst at a 2–10 keV flux level of  $\sim 4 \times 10^{-13} \text{ erg cm}^{-2} \text{ s}^{-1}$ . The equivalent spectral slope between the I-band and the X-ray is  $\alpha_{IX} = -0.7 \pm 0.1$ , a slope which is also consistent with the typical spectral indices of X-ray afterglows themselves:  $-1.0 < \alpha_X < -0.5$  (refs 19, 26, 27). Only the steeper optical spectrum of GRB971214 deviates from the empirical evidence that a single power law approximates the afterglow spectrum once the synchrotron break frequency passes below the optical band. With regard to afterglow theory, an intrinsic optical spectrum as steep as  $\alpha = -2.5$  is



**Figure 2** Photometry of GRB971214 taken from Table 1. The tentative J-band detection on December 16.52 is arbitrarily assigned an error bar of  $\pm 0.5$ . The I-band observation of December 31.46 yielded only an upper limit. Solid lines are power laws of slope  $-1.4$  drawn between observations taken with the same telescope and filter. Dashed lines indicate estimated uncertainties of  $\pm 0.2$  on the slope.



**Figure 3** Calculated spectrum of the optical afterglow. The photometry obtained by various telescopes on December 15.5, taken from Table 1, has been used to synthesize a contemporaneous spectrum of the optical afterglow 12 h after the burst. A power law of the form  $F_\nu \propto \nu^{-2.5}$  is drawn through the *VRI* points.

inconsistent with the simplest form of the external relativistic blast wave model<sup>28,29</sup>, which predicts that the flux density at a fixed observed frequency evolves as  $F_\nu(t) \propto \nu^\alpha t^\delta$ , where  $\delta = 3\alpha/2$ . The value of  $\delta$  implied by this relation and our measured spectral slope for GRB971214 is  $-3.75$ , much steeper than the observed decay of index  $-1.4 \pm 0.2$ .

We now estimate the amount of intrinsic extinction needed to de-redden the observed spectrum of GRB971214 to the colour of  $V - I = 0.9$ , which is characteristic of GRB970508 during its first five days<sup>22</sup>. The result depends on redshift and on the uncertain form of the ultraviolet extinction law. Assuming an average Galactic interstellar extinction law<sup>30</sup> at redshift  $z = 1$ ,  $E(B - V) \approx 0.25$  is required. The de-reddened I-band magnitude on December 15.5 in this case would be 20.2, closer to the peak fluxes of the previous optical afterglows. Similarly, for assumed redshift  $z = 2$ ,  $E(B - V) \approx 0.40$ , is required, together with de-reddened  $I = 18.6$ . For this plausible range of redshifts, substantial intrinsic extinction, corresponding to hydrogen column density in the range  $(1.5 - 2.3) \times 10^{21} \text{ cm}^{-2}$ , is thus implied. These values may be considered lower limits to the extent that the comparison object, GRB970508, could itself be somewhat reddened<sup>31</sup>.

It is likely that this extinction is due to interstellar dust in the host galaxy of the GRB rather than an intervening galaxy along the line of sight. Although there are two galaxies within 5 arcsec of the optical transient (see Fig. 1), it is unlikely that either contains the needed column density of material  $\geq 20 \text{ kpc}$  from its centre (assuming  $z \geq 0.2$ ,  $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ), which likewise disfavours either as the birthplace of the GRB progenitor. We therefore infer that a fainter host galaxy is present at the position of the transient. Preliminary indications from the Keck telescope are that the optical fading of GRB971214 ceased by 10 January 1998, leaving a co-incident object of  $R \approx 25.6 \text{ mag}$  which could be that host<sup>32</sup>. Candidate redshifts of 3.43 and 0.44 were suggested for the host based on Keck spectroscopy<sup>33</sup>.

Although the cosmological distance and extreme inferred energy,  $\sim 10^{52} \text{ erg}$  in  $\gamma$ -rays, seem well supported by the afterglow data (at least for GRB970508 at  $z \geq 0.835$ ; ref. 5), the actual source of that energy remains unknown. Widely discussed models for the underlying process in GRBs involve the merger of compact objects such as pairs of neutron stars or neutron-star/black-hole binaries. However, it is not clear that either of these events could channel  $\sim 10^{52} \text{ erg}$  into  $\gamma$ -rays, because the efficiency would have to be a few per cent<sup>34</sup>. For this reason, potentially more energetic reservoirs have been hypothesized; for example, a spinning black hole formed from the collapse of a massive, rotating star<sup>35</sup>. But it is unlikely that the energetics problem will be solved by theory alone. The magnitude of the problem needs to be quantified with many more redshift measurements from optical afterglows, which will determine the full distribution of GRB luminosities. The optical data are also needed to identify host galaxies, most importantly to determine how far the GRB progenitors can have travelled from the locations of their birth. This distance is an indicator of their total lifetimes, and therefore a potential restriction on the types of individual stars or stellar systems from which the GRB progenitors are drawn.  $\square$

Received 22 January; accepted 3 March 1998.

- Groot, P. J. *et al.* *IAU Circ.* No. 6588 (1997).
- Metzger, M. R. *et al.* *IAU Circ.* No. 6588 (1997).
- van Paradijs, J. *et al.* Transient optical emission from the error box of the  $\gamma$ -ray burst of 28 February 1997. *Nature* **386**, 686–689 (1997).
- Sahu, K. C. *et al.* The optical counterpart to the  $\gamma$ -ray burst GRB970228 observed using the Hubble Space Telescope. *Nature* **387**, 476–478 (1997).
- Metzger, M. R. *et al.* Spectral constraints on the redshift of the optical counterpart to the  $\gamma$ -ray burst of 8 May 1997. *Nature* **387**, 878–880 (1997).
- Groot, P. J. *et al.* *IAU Circ.* No. 6584 (1997).
- Bond, H. E. *et al.* *IAU Circ.* No. 6654 (1997).
- Djorgovski, S. G. *et al.* The optical counterpart to the  $\gamma$ -ray burst GRB970508. *Nature* **387**, 876–878 (1997).
- Sahu, K. C. *et al.* Observations of GRB 970228 and GRB 970508 and the neutron star merger model. *Astrophys. J.* **489**, L127–L131 (1997).
- Pian, E. *et al.* Hubble Space Telescope imaging of the optical transient associated with GRB 970508. *Astrophys. J.* **492**, L103–L106 (1998).

11. Pederson, H. *et al.* Evidence for diverse optical emission from gamma-ray burst sources. *Astrophys. J.* **496**, 311–315 (1998).
12. Heise, J. *et al.* *IAU Circ.* No. 6787 (1997).
13. Halpern, J., Thorstensen, J., Helfand, D. & Costa, E. *IAU Circ.* No. 6788 (1997).
14. Diercks, A. *et al.* *IAU Circ.* No. 6791 (1997).
15. Castander, F. J. *et al.* *IAU Circ.* No. 6791 (1997).
16. Rhoads, J. *IAU Circ.* No. 6793 (1997).
17. Tanvir, N., Wyse, R., Gilmore, G. & Corson, C. *IAU Circ.* No. 6796 (1997).
18. Landolt, A. U. UBVR photometric standard stars in the magnitude range  $11.5 < V < 16.0$  around the celestial equator. *Astron. J.* **104**, 340–371 (1992).
19. Costa, E. *et al.* Discovery of an X-ray afterglow associated with the  $\gamma$ -ray burst of 28 February 1997. *Nature* **387**, 783–785 (1997).
20. Antonelli, L. A. *et al.* *IAU Circ.* No. 6792 (1997).
21. Piro, L. *et al.* Evidence for late-time outburst of the X-ray afterglow of GB970508 from BeppoSAX. *Astron. Astrophys.* **331**, L41–L44 (1998).
22. Sokolov, V. V. *et al.* in *Proc. 4th Gamma-Ray Burst Symp.* (eds Meegan, C., Preece, R. & Koshut, T.) (AIP, New York, 1997).
23. Livio, M. *et al.* in *Proc. 4th Gamma-Ray Burst Symp.* (eds Meegan, C., Preece, R. & Koshut, T.) (AIP, New York, 1997).
24. Galama, T. *et al.* The decay of the optical emission from the  $\gamma$ -ray burst GRB970228. *Nature* **387**, 479–481 (1997).
25. Bessel, M. S. UBVR photometry. II—The Cousins VRI system, its temperature and absolute flux calibration, and relevance for two-dimensional photometry. *Publ. Astron. Soc. Pacif.* **91**, 589–607 (1979).
26. Piro, L. From the gamma-ray burst to its afterglow with BeppoSAX. *Bull. Am. Astron. Soc.* **29**, 1303 (1997).
27. Frontera, F. *et al.* Spectral properties of the prompt x-ray emission and afterglow from the gamma-ray burst of 1997 February 28. *Astrophys. J.* **493**, L67–L70 (1998).
28. Mészáros, P. & Rees, M. J. Optical and long-wavelength afterglow from gamma-ray bursts. *Astrophys. J.* **476**, 232–237 (1997).
29. Wijers, R. A. M., Rees, M. J. & Mészáros, P. Shocked by GRB 970228: the afterglow of a cosmological fireball. *Mon. R. Astron. Soc.* **288**, L51–L56 (1997).
30. Savage, B. D. & Mathis, J. S. Observed properties of interstellar dust. *Annu. Rev. Astron. Astrophys.* **17**, 73–111 (1979).
31. Reichart, D. E. The redshift of GRB 970508. *Astrophys. J.* **495**, L99–L102 (1998).
32. Kulkarni, S. R. *et al.* GCN Message No. 27 (1998).
33. Kulkarni, S. R., Adelberger, K. L., Bloom, J. S., Kundic, T. & Lubin, L. GCN Message No. 29 (1998).
34. Narayan, R., Paczyński, B. & Piran, T. Gamma-ray bursts as the death throes of massive binary stars. *Astrophys. J.* **395**, L83–L86 (1992).
35. Paczyński, B. Are gamma-ray bursts in star forming regions? *Astrophys. J.* **494**, L45–L48 (1998).
36. Monet, D. *et al.* USNO-SAI.0 (US Naval Observatory, Washington DC, 1996).

Correspondence and requests for materials should be addressed to J.P.H. (jules@astro.columbia.edu).

## The energetic afterglow of the $\gamma$ -ray burst of 14 December 1997

A. N. Ramaprakash<sup>\*†</sup>, S. R. Kulkarni<sup>\*</sup>, D. A. Frail<sup>‡</sup>, C. Koresko<sup>\*</sup>, M. Kuchner<sup>\*</sup>, R. Goodrich<sup>§</sup>, G. Neugebauer<sup>\*</sup>, T. Murphy<sup>\*</sup>, S. Eikenberry<sup>\*</sup>, J. S. Bloom<sup>\*</sup>, S. G. Djorgovski<sup>\*</sup>, E. Waxman<sup>||</sup>, F. Frontera<sup>¶</sup>, M. Feroci<sup>⊞</sup> & L. Nicastro<sup>\*\*</sup>

<sup>\*</sup>Palomar Observatory 105-24, California Institute of Technology, Pasadena, California 91125, USA

<sup>†</sup>Inter-University Centre for Astronomy and Astrophysics, Ganeshkhind, Pune 411 007, India

<sup>‡</sup>National Radio Astronomy Observatory, Socorro, New Mexico 87801, USA  
<sup>§</sup>W. M. Keck Observatory, 65-1120 Mamalahoa Highway, Kamuela, Hawaii 96743, USA

<sup>||</sup>Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540, USA

<sup>¶</sup>Istituto Tecn. Studio Rad. Extraterrestri, CNR, via Gobetti 101, Bologna I-40129, Italy

<sup>⊞</sup>Dipartimento di Fisica, Università Ferrara, via Paradiso 12, I-44100 Ferrara, Italy

<sup>\*\*</sup>Istituto di Astrofisica Spaziale, CNR, via Fosso del Cavaliere, Roma I-00133, Italy

<sup>\*\*</sup>Istituto Fisica Cosmica App. Info., CNR, via U. La Malfa 153, Palermo I-90146, Italy

The discovery of fading but relatively long-lived X-ray emission<sup>1</sup> accompanying  $\gamma$ -ray bursts has revolutionized the study of these objects. This ‘afterglow’ is most easily explained by models<sup>2–4</sup> similar to those describing supernovae, but with relativistic ejecta. And as with supernovae, afterglow measurements should in principle provide important constraints on burst properties,

permitting, for example, estimates of the amount of energy released, the geometry of the emitting surface and the density of the ambient medium. Here we report infrared observations of the fading optical transient<sup>5</sup> associated with the burst of 14 December 1997 (GRB971214; ref. 6). We detect a ‘break’ in the broad-band spectrum, as predicted by afterglow models, which constrains the total energy in the burst to be  $>10^{51}$  erg. Combining the fluence of optical afterglow with the redshift ( $z = 3.42$ ; ref. 7), we estimate that the energy released in the afterglow alone was  $2 \times 10^{51}$  erg. Estimates of afterglow energetics are less likely to be subject to geometric effects—such as beaming—that render uncertain estimates of the total burst energy, but it nevertheless appears from our measurements that  $\gamma$ -ray bursts may be much more energetic than the  $10^{51}$  erg usually assumed.

Our primary goal was to infer fundamental parameters of the burst from observations. As explained in detail below, one of the best diagnostics is a single broad-band (optical, near-infrared and radio) spectrum of the optical transient (OT) at a given time. Consequently we undertook observations in the near-infrared band and at radio wavelengths. Observations in the optical band began on the Keck II telescope as soon as we were notified of the localization by the Wide Field Camera<sup>6</sup> on board the X-ray satellite BeppoSAX. These observations are discussed in the companion Article<sup>7</sup>.

The near-infrared observations were initiated on the Keck I telescope only after the receipt of the tighter localization (90% confidence radius of 1 arcmin) resulting from observations of the X-ray afterglow<sup>8</sup> by the Narrow Field Instruments (NFI), also on board BeppoSAX. We used the Near Infrared Camera<sup>9</sup> (NIRC) with an InSb Hughes-SBRC array ( $256 \times 256$  pixels) and imaged in the K<sub>s</sub> band. The log of the observations is given in Table 1. Despite the tight NFI localization, the small field of view of NIRC ( $38 \text{ arcsec} \times 38 \text{ arcsec}$ ) required a large number of NIRC pointings to cover the error circle.

The OT is quite well detected in the December 15 data (Fig. 1). The K-band magnitude of the OT at epoch 1997 December 15.58 UT is found to be  $19.4 \pm 0.2 \text{ mag}$  (see Table 1). The same sequence of NIRC pointings was repeated the next night, but with a revised NFI localization which unfortunately did not include the OT<sup>5</sup>. We obtained considerably deeper observations in February (see Table 1). No object was detected at the position of the OT. The  $3\sigma$  upper limit to a point source obtained by combining the two February data sets is 22.5 mag. The limit for an extended source is weaker. The absence of an infrared source is not surprising given the power-law decay shown by the OT at optical frequencies<sup>7,10</sup>. However, Kulkarni

**Table 1 Log of K-band observations**

| Epoch (UT)      | Band           | Integration (s) | Magnitude          | Seeing (arcsec) | Observer |
|-----------------|----------------|-----------------|--------------------|-----------------|----------|
| 1997 Dec. 15.61 | K <sub>s</sub> | $9 \times 20$   | 19.4               | 0.7             | C.K.     |
| 1997 Dec. 16.66 | K <sub>s</sub> | $9 \times 20$   | (*)                | 0.7             | C.K.     |
| 1998 Feb. 07.46 | K              | 1,200           | $>21.5^{\dagger}$  | 0.4             | G.N.     |
| 1998 Feb. 10.56 | K              | 4,320 (c)       | $>22.1^{\ddagger}$ | 0.5             | R.G.     |

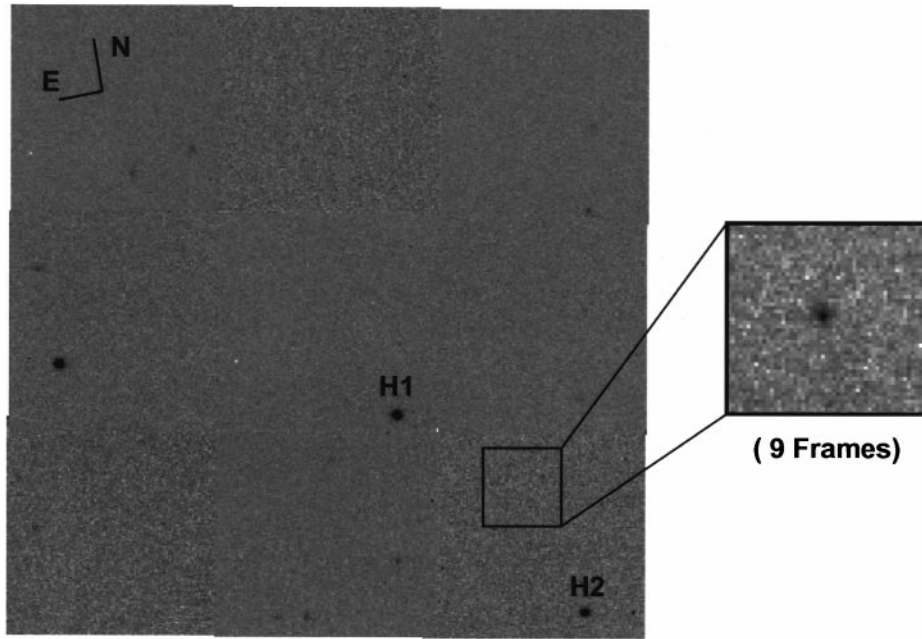
The K<sub>s</sub> band covers the range 1.99–2.32  $\mu\text{m}$ . The K-band filter covers the range 2.00–2.42  $\mu\text{m}$ . All quoted magnitudes have been transformed to the K band. The K-band magnitude of the OT was measured with respect to stars H1 and H2 (Fig. 1). One of us (G.N.) observed the field of OT on 1998 February 7 UT and the standard star SC9139 (E. Persson, personal communication) and derived  $K = 16.26 \pm 0.03$  for H2. This set the zero point for all of our K and K<sub>s</sub> observations. The derived K magnitude for H1 is  $15.31 \pm 0.03$ , which agrees very well with that expected from its BVR (ref. 26) and I (ref. 10) colours.

\*The mosaic did not include the position of the OT. See text.

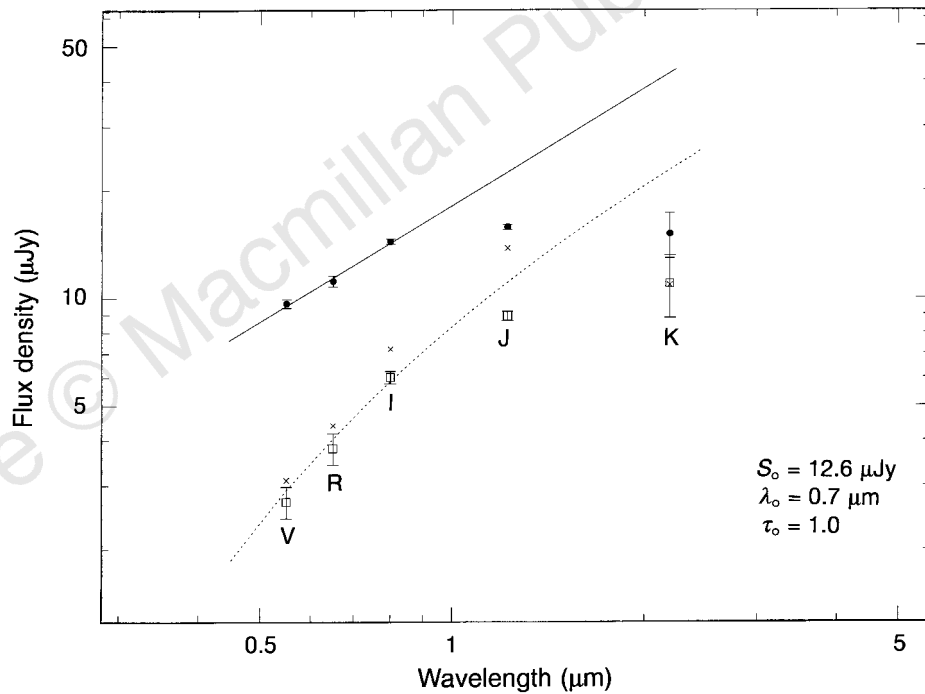
<sup>†</sup>The  $1\sigma$  limit for a beam of 1.5 arcsec diameter is  $K = 23.5 \text{ mag}$ . The limit quoted in the table is the  $3\sigma$  limit for a point source. This was obtained by processing simulated point sources with a similar point-spread function as H2, and then using IRAF task DAOFIND to detect the object.

<sup>‡</sup>We used a nine-position dither pattern with exposure time of 30 s at each position; 16 of these dither patterns were summed together for a total exposure time of 72 min. Each individual frame within a dither pattern was shifted and added according to the commanded telescope offsets; the scarcity of stars in the field precluded using stars for this ‘shift-and-add’ procedure. The 16 mosaics thus formed were shifted and added using the bright star at the edge of the field. The  $3\sigma$  upper limit that is quoted was derived in a manner similar to that in the above footnote.





**Figure 1**  $K_s$ -band image of the field of the optical transient (OT) of GRB971214 obtained using NIRC. The figure on the left is a single  $3 \times 3$  mosaic. The mosaic was formed by offsetting each NIRC frame by 30 arcsec in either one or both of the cardinal coordinates so as to form a square of side 98 arcsec. Nine such mosaics were obtained with their centres on a  $3 \times 3$  grid with a grid spacing of 10 arcsec. Two nearby bright reference stars (H1 and H2) are marked. The position of the OT is at the centre of the inset  $15'' \times 15''$  square. It is not visible in a single mosaic but is readily detected when all the nine mosaics are summed; see inset on the right side. The data were analysed in the usual manner: a 'super-sky' frame was obtained by combining all the frames. This super-skyflat frame was subsequently used to flat-correct each of the mosaics. The software package IRAF (Image Reduction and Analysis Facility; distributed by the National Optical Astronomy Observatories) was used to process the data.



**Figure 2** The optical/near-infrared broad-band spectrum of GRB971214 at epoch 1997 December 15.58 UT, 0.58 d after the burst. The crosses with no error bars are the original data taken from the literature and this Letter with  $V = 22.65 \pm 0.1$  mag (0.55  $\mu\text{m}$ ; ref. 10) at 1997 December 15.51 UT,  $R = 22.1 \pm 0.1$  mag (0.65  $\mu\text{m}$ ; ref. 19) at 1997 December 15.51 UT,  $I = 21.3 \pm 0.04$  mag (0.8  $\mu\text{m}$ ; ref. 10) at 1997 December 15.50 UT,  $J = 20.27 \pm 0.03$  mag (1.22  $\mu\text{m}$ ; ref. 20) at 1997 December 15.44 UT,  $K = 19.4 \pm 0.2$  mag (2.19  $\mu\text{m}$ , this Letter) at 1997 December 15.58 UT. The observed magnitudes were converted to spectral flux densities assuming the following fluxes for zero mag: 3,590 Jy (V; ref. 27), 3,020 Jy (R; ref. 27), 2,380 Jy (I; ref. 27), 1,570 Jy (J; ref. 28) and 636 Jy (K; ref. 28). The observed fluxes were extrapolated to the same epoch as the K-band measurement, 1997 December 15.58 UT, using the same temporal slope as the I-band light curve:  $I(t) = 22.22 + 3.09 \log(t - t_0)$ , where  $t_0 = 1997$  December 14.9727 UT. A power-law model fitted to these extrapolated VRI points (open points) yields a power-law index of  $2.13 \pm 0.04$ ; the power-law fit is shown by the dotted line. As explained in the text, in the framework of the afterglow models the spectrum of the decaying afterglow is expected to be a power law with index  $\alpha = a/b$ ; here  $a$  is the power-law index of the afterglow observed at any one frequency and  $b = 3/2$  (see equation (2)). From I-band observations discussed in ref. 7 we note that  $a = 1.2 \pm 0.2$  and thus the

expected  $\alpha$  is  $0.8 \pm 0.13$ . To reconcile the observations with theory we invoke extinction by dust between us and the OT. To this end we used a two-component model consisting of a power law,  $S_0(\lambda/\lambda_0)^\alpha$ , and an exponential absorption,  $\exp(-\lambda_0\tau_0/\lambda)$ . The first component is the expected spectral behaviour of the afterglow from GRB971214;  $\alpha = 0.8$ . The second component is the 'extinction law'; we assume the optical depth from extinction is  $\propto \lambda^{-1}$ . The parameters  $\tau_0$  and  $\lambda_0$  are degenerate but are displayed as such for physical clarity. Only the VRI fluxes were used in the model fit as the OT is known to undergo a power-law decay in these bands and thus we have every expectation that the underlying spectrum is a power-law (see above). The best-fit parameters (solid line) are  $\alpha = 0.8$  (by fiat),  $\tau_0\lambda_0 = 0.7 \mu\text{m}$  and  $S_0 \approx 13 \mu\text{Jy}$ . A final set of points (filled circles) shows the flux density when data are corrected for the extinction law derived above. The corrected J and K fluxes do depend on the extrapolation of the extinction law to these wavelengths. Nonetheless, as can be seen from the figure, the VRI fluxes are well fitted by the power-law dependence (solid line) expected for the afterglow but there is clear evidence of a spectral turnover or flattening in the J and K bands, indicating that the  $\nu F_\nu$  spectrum peaked at around 1  $\mu\text{m}$  at the epoch of the K-band observations.

*et al.*<sup>7</sup> report the detection of a galaxy with R-band magnitude 25.6 at the position of the OT and suggest that this is the host of the  $\gamma$ -ray burst (GRB). This galaxy is found to have a redshift<sup>7</sup> of 3.4 and a spectrum similar to star-forming galaxies<sup>11</sup> at such redshifts. The  $R-K$  colour for such galaxies is found to be 2.3 to 3.0 mag (K. L. Adelberger, personal communication). This is consistent with our K-band limit.

Observations at 8.46 GHz and centred on the position of the fading X-ray source<sup>8</sup> were conducted with the Very Large Array (VLA) on UT 1997 December 15.67, 16.34, 18.33, 20.28 and UT 1998 January 03.38. The r.m.s. noise in the images obtained from each of the observations ranged from 50 to 11  $\mu$ Jy per beam. The r.m.s. noise in the summed image is 7.8  $\mu$ Jy per beam. No radio counterpart to the OT was seen in any of these images.

The lower-energy emission—X-ray, optical, infrared and radio—that follows some of the bursts is referred to as the afterglow emission. The simplest afterglow model<sup>2–4</sup> posits an instantaneous release of energy ( $E_0$ ) and “ejecta” (mass,  $M_B$ ) in a homogeneous medium (particle density,  $n_0$ ). The ejecta “sweep up” the ambient gas and shock-accelerate electrons to a power-law energy distribution. The energy fractions in electrons and post-shock magnetic fields are assumed to be close to the equipartition values. These three assumptions are routinely used to explain shocks at other astrophysical sites such as supernovae and jets in active galactic nuclei. The afterglow is due to synchrotron emission by the relativistic electrons. This model seems to explain quite well multi-wavelength observations of GRB970228 (ref. 12) and GRB970508 (ref. 13–15).

A generic prediction of such models is that the spectrum of the afterglow has the following form:

$$f(\nu) = S_0(\nu/\nu_b)^{-\alpha}, \quad \nu > \nu_b, \quad (1a)$$

$$= S_0(\nu/\nu_b)^{-\beta}, \quad \nu < \nu_b \quad (1b)$$

The power-law exponent  $\alpha$  is related to the power-law distribution of the electron energies. The exponent  $\beta$  depends weakly on the low-energy electron tail of the shocked electrons; a robust value is 1/3 (ref. 16).  $\nu_b$  is the “break” frequency and is expected to evolve as follows<sup>16</sup>:

$$\nu_b = 2 \times 10^{14} \left( \frac{1+z}{2} \right)^{1/2} (\xi_e/0.2)^2 (\xi_B/0.1)^{1/2} E_{52}^{1/2} t^{-b} \text{ Hz} \quad (2)$$

with  $b = 3/2$ . Here  $\xi_e$  and  $\xi_B$  are fractions which measure the degree to which the electrons and magnetic fields are in equipartition with respect to the shocked protons. The total energy of the GRB is  $E_0 = 10^{52} E_{52}$  erg,  $z$  is the redshift of the burst, and  $t$  is the time (in days) after the burst.

Two assumptions are implicit in the above discussion. (1) The shock is non-radiative, that is, the cooling time for the bulk of the synchrotron emitting electrons is larger than the ‘dynamical’ time over which the relativistic blast wave evolves. According to Waxman *et al.*<sup>17</sup>, the radio and the optical observations of GRB970508 strongly support this assumption. (2) At any given time, electrons producing high-frequency radiation cool more rapidly than those producing low-frequency radiation. This gives rise to another break at high frequencies. Sari *et al.*<sup>18</sup> suggest that the late time optical afterglow of GRB970508 is in this regime. We address this concern below.

From equations (1) and (2) it may be shown that at a fixed frequency  $\nu > \nu_b$ , the flux decays as  $S_\nu(t) \propto t^{-a}$ , where  $a = b\alpha$ . Thus  $\alpha$  can either be measured directly from the spectrum or from observations of the light curve in a fixed band. A measurement of a spectrum across a large range in frequency (broad-band) at an epoch, say  $t_*$  (here and below, all times are measured with respect to the GRB event) yields  $S_0$ ,  $\nu_* \equiv \nu_b(t_*)$ ,  $\alpha$  and  $\beta$ . As can be seen from

equations (1) and (2), once these parameters are known the subsequent evolution of the afterglow emission is fully determined.

In Fig. 2 we show the observed fluxes of the OT across the optical<sup>7,10,19</sup> bands, a near-infrared band<sup>20</sup> and our own K-band measurement; all these measurements were made around 1997 December 15.5 UT. As can be seen from this figure, the broad-band spectrum of the OT has roughly the shape described by equation (1). However, there is one notable discrepancy. From equations (1) and (2) we predict that  $\alpha = a/b = 0.8 \pm 0.15$ . Instead, the power-law index for the observed VRI fluxes is  $2.13 \pm 0.04$  (see Fig. 2). This discrepancy has already been noted<sup>10,21</sup> and attributed to extinction by intervening dust. Dust preferentially extinguishes (absorbs and scatters) shorter-wavelength light, steepening the curve.

Correction for extinction by dust requires knowledge of the wavelength dependence of this process (the extinction law) in the host galaxy of the OT. The data for other galaxies, even the two nearest (the Large and the Small Magellanic Cloud), are scanty. We assume a simple dust extinction law (extinction optical depth proportional to inverse wavelength) and deduce the extinction from the observed VRI measurements (see Fig. 2). As can be seen from this figure, the observed, as well as the extinction-corrected, J- and K-band fluxes are well below the extrapolation of the model. One interpretation is that extinction correction has a more complicated frequency dependence. We prefer the simpler interpretation that the optical–infrared broad-band spectrum shows the two-component structure described by equation (1). This interpretation is fully consistent with our  $3\sigma$  upper limit on the 8.46-GHz flux discussed earlier. Accepting this interpretation, we find that the break frequency is  $\nu_* \approx 3 \times 10^{14}$  Hz (corresponding to a wavelength of 1  $\mu$ m) at a post-burst time of  $t_* = 0.58$  d. We note that the significant change in slope across 1  $\mu$ m is inconsistent with the idea that this break is due to the electron cooling discussed above<sup>18</sup>.

The determination of  $\nu_*$  allows us to determine the energy of the GRB. Substituting  $\nu_*$  and  $t_*$  into equation (2) we find:

$$E_{52} = 0.44 \left( \frac{2}{1+z} \right) \left( \frac{\xi_e}{0.2} \right)^{-4} \left( \frac{\xi_B}{0.1} \right)^{-1} \quad (3)$$

In this equation the equipartition factors are set close to their maximum value (their sum clearly cannot exceed the contribution from protons). Thus, independent of the ambient density,  $n_0$ , we note that  $E_{52} > 0.1$ . The dependence on the redshift is weak and is stronger for smaller values of  $z$ .

A notable prediction of this hypothesis is that the OT should be detected in the radio band with a flux of  $\sim 13$   $\mu$ Jy. For GRB970508, the radio flux rose rapidly and reached the peak value within a few weeks<sup>17,22</sup>. Thus transparency at radio wavelengths is probably not an issue. We look forward to further radio observations of GRB 971214.

The fluence in the afterglow, integrated up to photon energy  $E_m = h\nu_m$  can be shown to be  $F_A = [a/(a-1)][b/(b-1)]S_0 t_* \nu_* \times (\nu_m/\nu_*)^{(b-1)/b}$ . We obtain  $F_A = 7 \times 10^{-8} (E_m/10 \text{ eV})^{1/3} \text{ erg cm}^{-2}$ . The X-ray emission dominates the energy budget of the afterglow. However, at early times (when the afterglow is mainly X-ray emission) our assumption of non-radiative shock may not be valid. For this reason we restrict the estimation of the fluence to lower photon energies.

Kulkarni *et al.*<sup>7</sup> have proposed that the OT lies in a galaxy with redshift  $z = 3.418$ . The luminosity distance ( $d_L$ ) for a Hubble constant  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , and  $\Omega = 0.3$ , is  $9.7 \times 10^{28} \text{ cm}$ . The luminosity in the afterglow, assuming spherical symmetry, is  $E_A(E_m < 10 \text{ eV}) = 4\pi d_L^2 (1+z)^{-1} F_A = 1.9 \times 10^{51} \text{ erg}$ . The  $\gamma$ -ray fluence as observed by the Gamma-Ray Burst Monitor (GRBM) on board BeppoSAX is  $(0.9 \pm 0.09) \times 10^{-5} \text{ erg cm}^{-2}$  (above 40 keV); the fluence<sup>23</sup> as measured by the Compton Gamma-ray Observatory (above 20 keV) is  $(1.09 \pm 0.07) \times 10^{-5} \text{ erg cm}^{-2}$ .

Thus the inferred isotropic energy loss in  $\gamma$ -rays is  $E_G \sim 3 \times 10^{53}$  erg. This is remarkably coincident with the gravitational binding energy released during the merger of two neutron stars<sup>24</sup>.

Even though  $E_G$  is much greater than  $E_A$ , the afterglow measurements pose a clear problem to GRB models. At the time the  $\gamma$ -rays are emitted, the shocked gas is moving with very large Lorentz factor ( $\Gamma$ ). The initial  $\Gamma$  is assumed to be  $10^2$ – $10^3$  and expected to evolve as  $t^{-3/8}$ . Due to relativistic beaming, only a small portion, solid angle  $\sim \Gamma^{-2}$ , of the fireball is visible to us. Hence we remain ignorant of the true shape of the fireball. The estimate of  $E_G$  can be reduced by assuming that the emitting surface is a narrow jet directed towards the observer. However,  $\Gamma$  is smaller during the optical afterglow emission than during the  $\gamma$ -ray emission. Relativistic beaming is less of a limitation, and the full extent of the fireball is gradually revealed to the observer. The monotonic power-law decay at optical wavelengths seen in the OTs of previous transients<sup>13–15</sup> is not consistent with the narrow jet-like emitting surfaces. We recognize that more complicated models (in which energy is injected non-spherically) could account for such monotonic decays whilst having non-spherical geometry. However, the main issue that we discuss here is the inferred energetics. It is by no means clear that the energy requirements inferred from afterglow observations (when the Lorentz factor is small) is any lower than the simplest case considered here.

According to the models discussed here, the afterglow emission is supposed to arise from non-radiative shocks; that is, the efficiency of the shock in converting  $E_0$  (the total energy in the GRB event) to afterglow radiation is low. Thus,  $E_0$  is significantly larger than the  $E_A$  reported above. This finding is in direct contrast to the assumptions made by most practitioners in this field, namely  $E_0 = 10^{51}$  erg. As an aside we note that equation (1) predicts that the break frequency should evolve with time as  $t^{-3/2}$ . Measurement of the break frequency at successive epochs for future bursts offers the possibility of directly confirming or refuting the non-radiative shock assumption.

The evidence presented here favours GRB models which produce energy vastly in excess of  $10^{51}$  erg; see for example, refs 24, 25, 29. The afterglow emission is similar in nature to the emission from supernovae but is more energetic by two orders of magnitude. Following the naming sequence, nova and supernova, it is only appropriate to refer to GRB afterglow as hypernova<sup>29</sup>.  $\square$

Received 18 March; accepted 14 April 1998.

- Costa, E. *et al.* Discovery of an X-ray afterglow associated with the  $\gamma$ -ray burst of 28 February 1997. *Nature* **387**, 783–785 (1997).
- Mészáros, P. & Rees, M. J. Optical and long-wavelength afterglow from gamma-ray bursts. *Astrophys. J.* **476**, 232–237 (1997).
- Vietri, M. The afterglow of gamma-ray bursts: the cases of GRB 970228 and GRB 970508. *Astrophys. J.* **488**, 232–237 (1997).
- Waxman, E. Gamma-ray-burst afterglow: supporting the cosmological fireball model, constraining parameters, and making prediction. *Astrophys. J.* **485**, L5–L8 (1997).
- Halpern, J., Thorstensen, J., Helfand, D. & Costa, E. *IAU Circ. No.* 6788 (1997).
- Heise, J. *et al.* *IAU Circ. No.* 6787 (1997).
- Kulkarni, S. R. *et al.* Identification of a host galaxy at redshift  $z = 3.42$  for the  $\gamma$ -ray burst of 14 December 1997. *Nature* **393**, 35–39 (1998).
- Antonelli, L. A. *et al.* *IAU Circ. No.* 6792 (1997).
- Matthews, K. & Soifer, B. T. in *Infrared Astronomy with Arrays* (ed. McLean, L.) 239–246 (Kluwer, Dordrecht, 1994).
- Halpern, J. P., Thorstensen, J. R., Helfand, D. J. & Costa, E. Optical afterglow of the  $\gamma$ -ray burst of 14 December 1997. *Nature* **393**, 41–43 (1998).
- Steidel, C. C., Giallisco, M., Pettini, M., Dickinson, M. & Adelberger, K. Spectroscopic confirmation of a population of normal star-forming galaxies at redshifts  $z > 3$ . *Astrophys. J.* **462**, L17–L21 (1996).
- Wijers, R. A. M. J., Rees, M. J. & Mészáros, P. Shocked by GRB 970228: the afterglow of a cosmological fireball. *Mon. Not. R. Astron. Soc.* **288**, L51–L56 (1997).
- Sokolov, V. V. *et al.* BVRC photometry of GRB 970508 optical remnant: May–August, 1997. *Astron. Astrophys.* (in the press); also preprint <http://xxx.lanl.gov, astro-ph/9802341> (1998).
- Pedersen, H. *et al.* Evidence for diverse optical emission from gamma-ray burst sources. *Astrophys. J.* **496**, 311–315 (1998).
- Galama, T. J. *et al.* Optical follow-up of GRB 970508. *Astrophys. J.* (in the press); also preprint <http://xxx.lanl.gov, astro-ph/9802160> (1998).
- Waxman, E. Gamma-ray burst afterglow: confirming the cosmological fireball model. *Astrophys. J.* **489**, L33–L35 (1997).
- Waxman, E., Kulkarni, S. R. & Frail, D. A. Implications of the radio afterglow from the gamma-ray burst of May 8, 1997. *Astrophys. J.* **497**, 288–293 (1998).
- Sari, R., Piran, T. & Narayan, R. Spectra and light curves of gamma-ray burst afterglows. *Astrophys. J.* (in the press).
- Diercks, A. *et al.* *IAU Circ. No.* 6791 (1997).
- Tanvir, N., Wyse, R., Gilmore, G. & Corson, C. *IAU Circ. No.* 6796 (1997).

- Reichert, D. E. The host galaxy of GRB 971214. *Astrophys. J.* (submitted); preprint <http://xxx.lanl.gov, astro-ph/9801139> (1998).
- Frail, D. A., Kulkarni, S. R., Nicastro, L., Feroci, M. & Taylor, G. B. The radio afterglow from the  $\gamma$ -ray burst of 8 May 1997. *Nature* **389**, 261–263 (1997).
- Kippen, R. M. *et al.* *IAU Circ. No.* 6789 (1997).
- Narayan, R., Paczynski, B. & Piran, T. Gamma-ray bursts as the death throes of massive binary stars. *Astrophys. J.* **395**, L83–L86 (1992).
- Rees, M. J. & Mészáros, P. Refreshed shocks and afterglow longevity in gamma-ray bursts. *Astrophys. J.* **496**, L1–L4 (1998).
- Henden, A. A., Luginbuhl, C. B. & Vrba, F. J. *GCN Note No.* 0.16 (1997); <http://gcns.gsfc.nasa.gov/gcn/gcn3/>.
- Fukugita, M., Shimasaku, K. & Ichikawa, T. Galaxy colors in various photometric band systems. *Publ. Astron. Soc. Pacif.* **107**, 945–958 (1995).
- Bessell, M. S. in *Stellar Photometry—Current Techniques and Future Developments* (eds Butler, C. J. & Elliot, I.) 22–39 (IAU Colloq. 136, Cambridge Univ. Press, 1992).
- Paczynski, B. Are gamma-ray bursts in star-forming regions? *Astrophys. J.* **494**, L45–L48 (1998).

**Acknowledgements.** The W. M. Keck Observatory is operated by the California Association for Research in Astronomy, a scientific partnership among California Institute of Technology, the University of California and NASA. It was made possible by financial support from the W. M. Keck Foundation. The VLA is a facility of the NSF operated under cooperative agreement by associated Universities, Inc. We thank W. Sargent, Director of Palomar Observatory, F. Chaffee, Director of the Keck Observatory and our colleagues for their continued support of our GRB program. S.R.K.'s research is supported by the NSF and NASA. S.G.D. acknowledges partial support from the Bressler Foundation. A.N.R. thanks the International Astronomical Union for a travel grant.

Correspondence and requests for materials should be addressed to S.R.K. (e-mail: [srk@surya.caltech.edu](mailto:srk@surya.caltech.edu)).

## Solid hydrogen at 342 GPa: no evidence for an alkali metal

Chandrabhas Narayana\*, Huan Luo\*†, Jon Orloff‡ & Arthur L. Ruoff\*

\* Department of Materials Science and Engineering, Cornell University, Ithaca, New York 14853, USA

‡ Institute for Plasma Research, University of Maryland, College Park, Maryland 20742-3511, USA

Solid hydrogen, an electrical insulator, is predicted to become an alkali metal under extreme compression, although controversy surrounds the pressure required to achieve this<sup>1–3</sup>. The electrical conductivity of hydrogen as a function of pressure and temperature is of both fundamental and practical interest—metallic hydrogen may be of relevance to planetary interiors<sup>4</sup>, and has been suggested as a potential high-temperature superconductor<sup>5</sup>. Calculations<sup>1,2</sup> suggest that depairing (destruction of the molecular bond) should occur around 340 GPa, accompanied by the formation of an alkali metal at this pressure<sup>1</sup>, or at substantially higher pressures<sup>2,3</sup>. Here we report that solid hydrogen does not become an alkali metal at pressures of up to  $342 \pm 10$  GPa, achieved using a diamond anvil cell. This pressure (which is almost comparable to that at the centre of the Earth) significantly exceeds those reached in earlier experiments—216 GPa (ref. 6) and 191 GPa (ref. 7)—at which hydrogen was found to be non-metallic. The failure of solid hydrogen to become an alkali metal at the extreme pressures reported here has implications for our current theoretical understanding of the solid-state phase.

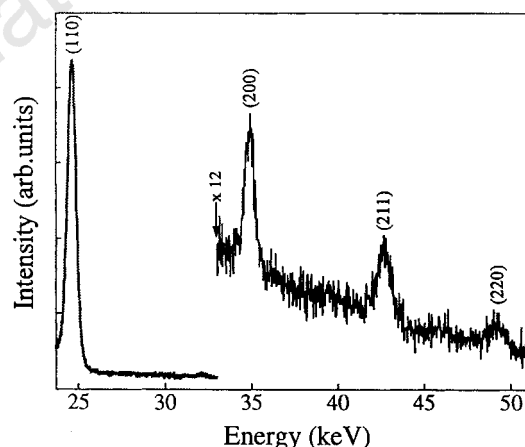
Several solids that are electrical insulators at atmospheric pressure have been made metallic at high pressure at 300 K (for example, BaSe<sup>8</sup>, xenon<sup>9,10</sup>, sulphur<sup>11</sup> and oxygen<sup>12</sup>) in the diamond anvil cell<sup>13</sup>. But the inability<sup>14–21</sup> to produce robust evidence for metallization of solid hydrogen by either optical or conductivity measurements remains<sup>22</sup>, although dynamic experiments<sup>23</sup> have provided evidence that liquid hydrogen becomes an electrical conductor at 4,000 K and 140 GPa. Calculations<sup>2</sup> on solid hydrogen suggest that a transition to a diamond-like cubic atomic phase (a gapless semiconductor or a semimetal) occurs around 340 GPa and that this should be transformed to a metallic simple cubic phase at 650 GPa and then to a metallic body-centred cubic phase around 900 GPa. Both of these

† Present address: Applied Materials, San Jose, California 95100, USA.



cubic phases have one free electron per atom. Accurately performing such experiments on solid hydrogen into the 340-GPa range and above has been called (H. D. Drickamer, personal communication) the most difficult experiment in solid-state science.

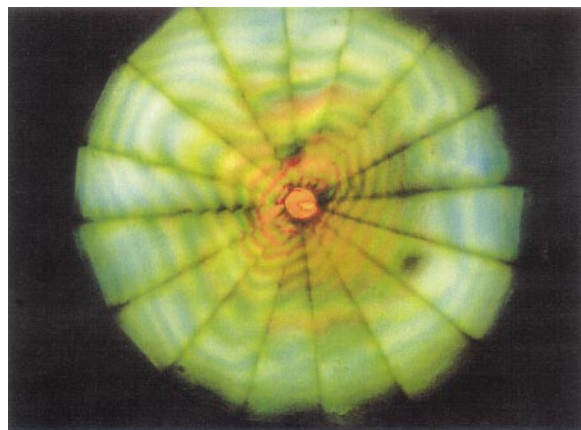
From 15 pairs of diamonds that we used to achieve megabar pressures on hydrogen, one pair reached a pressure of 255 GPa and one reached 275 GPa. In both cases, the diamonds had 35- $\mu\text{m}$  tips. To achieve higher pressure a pair of diamonds with 20- $\mu\text{m}$  tips and 8° bevels was used<sup>24,25</sup>, with a gasket pre-indented to a thickness of 10  $\mu\text{m}$ . The experiment which achieved 342 GPa used techniques crucially different from those used by other experimenters on hydrogen, as follows. (1) The tips were much smaller (20  $\mu\text{m}$ ). It was with such small tips that pressures of over 500 GPa were obtained on several metals<sup>25</sup>. (2) The gasket was tungsten. Unlike the case for rhenium, no dissolution of hydrogen into tungsten and no evidence of hydride formation was found by the present authors in two experiments: a pressure of 40 GPa for 30 d, and 221 GPa for 24 h. No decrease in pressure was detected, and no change in sample area was found. It is important to avoid nascent hydrogen which attacks the diamond and which is produced by hydride formers such as rhenium. (3) The sample holes were made with no burrs: an important factor, as burrs can lead to instabilities. The holes were centred  $\pm 0.25 \mu\text{m}$  on both 20- $\mu\text{m}$ -diameter indented flats; alignment is therefore maintained during pre-indenting and hence also in loading. The initial hole diameter was 11  $\mu\text{m}$ , so that the gasket extended over about two-thirds of the 20- $\mu\text{m}$ -diameter tip area initially and over 90% of the tip area at high pressures (the sample area decreases initially with increasing pressure, and at 275–342 GPa it was 0.29 of the original sample area). The critical importance of an initial sample radius which is a small fraction of the tip radius is discussed later. (4) The sample was brought to  $\sim 50$  GPa (from 0.204 GPa) in less than 1 min (it was still circular but with a smaller radius, as expected), and to 270 GPa in less than 30 min. This decreases the possibility of hydrogen attack on the diamond. (5) The lattice parameter (and hence the stress) in the tungsten gasket at the tungsten-sample interface was measured. The stress was found to be huge compared to the tungsten yield strength (and even to the yield strength of perfect tungsten crystals). This ensures that hydrogen was present at the measured pressure in the sample cavity. Measurement of the Raman shift of the vibron of hydrogen also showed that hydrogen is present<sup>26</sup>.



**Figure 1** An energy dispersive X-ray diffraction pattern of tungsten adjacent to the hydrogen sample at a hydrogen pressure of 342 GPa.  $E_d = 47.531 \text{ keV } \text{\AA}$ . Here  $E$  is the energy of a diffraction peak from a plane whose interplanar spacing is  $d$ . The energy peaks, fitted by gaussians, are located at 24.620, 34.904, 42.714 and 49.252 keV. The weighted average of the lattice parameters gives  $a_s = 2.7301 \text{ \AA}$  which gives an apparent pressure  $P_s = 322.2 \text{ GPa}$  (which must be corrected) when  $a_s$  is substituted in the isotherm. The value of  $a_0$  used was  $3.1653 \text{ \AA}$ .

Such measurements are routine below 150 GPa. However, at pressures above 150 GPa the hydrogen stiffens and becomes very non-hydrostatic. This causes line broadening and a substantial decrease in the signal to noise ratio. Our laser source was the 5,145- $\text{\AA}$  line of argon. This radiation causes substantial luminescence from the diamond at high pressure, further decreasing the signal to noise ratio. In an attempt to make Raman measurements above 150 GPa with our available equipment, a signal was obtained from the sample and then a reference signal was obtained from the gasket away from the sample hole: the apparent Raman signal was obtained by difference. This procedure could involve errors due to the curvature of the diamonds and due to the refractive-index difference of the two paths as well as in the way zero references are made, so we consider these results as preliminary.

All static high-pressure measurements above  $\sim 7$  GPa currently have as their basis the use of X-ray measurements of a marker material such as Mo, W, Pt, Cu, Ag or Au. An isotherm at (for example) 300 K has been derived for these materials from a measured Hugoniot,  $P_H(V/V_0)$ , using thermodynamics and statistical mechanics for the thermal pressure correction (which is least for W and then Mo). Here  $P_H$  is the Hugoniot pressure in a shock-experiment when the material is compressed from a volume  $V_0$  to  $V$ . Figure 1 shows the diffraction pattern of tungsten for a pressure of 342 GPa in the hydrogen at the gasket-sample interface. The apparent pressure,  $P_s$ , was obtained from the ratio of  $a/a_0$  or  $V/V_0 = 0.6446$ . Here  $a_0$  is the initial lattice parameter and  $a$  is the lattice parameter at pressure. The 300 K isotherm used a Birch<sup>27</sup> two-parameter fit ( $B_0 = 311.8 \text{ GPa}$  and  $B'_0 = 3.826$ ) to the data of McQueen *et al.*<sup>28</sup> which has been confirmed to 410 GPa (R. S. Hixson, private communication). Calculations (A.L.R., C. O. Rodriguez and N. E. Christensen, manuscript in preparation) of  $P(V/V_0)$  for W to 600 GPa show excellent agreement with experiment. Mo has also been studied in detail by the same group of experimentalists<sup>29,30</sup> and theorists<sup>31</sup>, again with excellent agreement which adds further credence to the isotherm used for W. It is necessary to make a correction to  $P_s$  because there is a non-hydrostatic state of stress in the tungsten (at the sample interface) for which the three principal stresses are  $P + \sigma_o$ ,  $P$  and  $P$ . Here  $P$  is the pressure in the hydrogen at the interface and  $\sigma_o$  is the yield stress of the tungsten at a mean normal stress  $\sigma_n = P + (\sigma_o/3)$ . The  $P + \sigma_o$  stress is along the axes of the diamonds and along the beam



**Figure 2** Reflectance and transmission micrograph of the sample and tip region, and birefringence of the diamond, as viewed with polarized light at 327 GPa. The nearly circular yellow area represents the flat tip (20- $\mu\text{m}$  diameter). Earlier studies<sup>35</sup> of the absorption edge of diamond to 421 GPa show that at 327 GPa, absorption by the diamonds would make incident white light appear yellow in transmission.

direction.  $\sigma_n$  is related to  $P_\sigma$  by<sup>32</sup>  $\sigma_n = P_\sigma + (BF/2G)\sigma_{00}(1 - 3\sin^2\theta)$  where  $B$  is the bulk modulus,  $G$  is the shear modulus,  $F$  is the pressure strengthening factor<sup>31,33</sup> (all at the mean normal stress,  $\sigma_n$ ),  $\sigma_{00}$  is the yield strength of the heavily strain-hardened tungsten when the pressure is removed, and  $\theta$  is the angle between the diffracting planes and the X-ray beam (whose direction is along the axes of the diamond anvils). How the present authors obtain  $F$  is given in ref. 31. For tungsten (A.L.R., C. O. Rodriguez and N. E. Christensen, manuscript in preparation)  $BF/G = 1.612 + 0.0185P - (3.70 \times 10^{-6})P^2$  for  $P < 600$  GPa, and  $F = 1 + 0.00859P - (4.31 \times 10^{-6})P^2$  for  $P < 600$  GPa. For heavily strain-hardened tungsten subsequently reduced from 250  $\mu\text{m}$  to several micrometres, the yield stress is<sup>32</sup>  $\sigma_{00} = 8$  GPa. (As  $\sigma_o = F\sigma_{00}$ , it follows that at  $P = 200$  GPa,  $\sigma_o = 20.4$  GPa in close agreement with a measured value<sup>34</sup> at 200 GPa.) An iterative process using the expression for  $BF/G$  and beginning with  $\sigma_n = P_\sigma$  leads to  $\sigma_n = 351.2$  GPa in the present case, and hence to  $P = 342$  GPa.

In the geometry used here, it has been observed in one case that pressure measured as described dropped slowly from 203 GPa after additional turning of the loading screw. Examination showed that the sample cavity had decreased in diameter and that a crack had occurred at the centre of the diamond. The measured pressure continued to decrease as the cavity diameter decreased and as the crack in the diamond enlarged, until finally the gasket hole was

closed. With the present geometry the measured stress decreases and the cavity closes when the hydrogen leaks away.

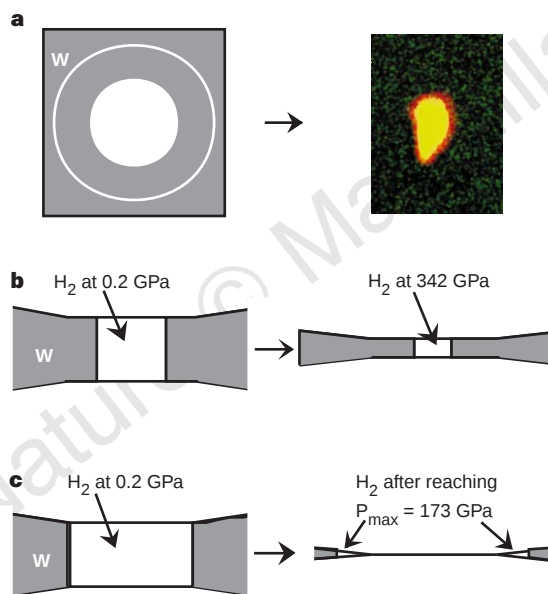
Figure 2 shows a combined reflection and transmission micrograph showing an approximately elliptical sample, a nearly circular tip and a birefringence pattern from the diamond at  $P = 327$  GPa. On the next load the authors obtained  $P = 342$  GPa. This pressure is more than 50% larger than previously reached on hydrogen, measured by acceptable methods by other workers<sup>6,7</sup>. After the pressure had been measured and the sample observed optically at 342 GPa (there was no perceptible change from that at 327 GPa), the torque on the loading bolt of the pressure cell was increased by an amount which corresponded in the previous increment to a further 15 GPa increase. The sample was observed and found to be still transparent.

Figure 3 shows the transmission of the incident white light through the sample at 327 GPa. The transmitted light appears yellow owing to absorption by the diamonds<sup>35</sup>. The transmission at 342 GPa appears unchanged from that at 327 GPa, as does the transmission when the torque was increased above the torque when the pressure was 342 GPa. There was no measurable change in the shape or size of the sample area between 275 and 342 GPa. With the sample in view in Fig. 3, it is useful to discuss the X-ray apertures used. These were  $6.0 \times 5.5 \mu\text{m}$  and were produced by masking and ion etching. The ratio of photon fluence at a fixed beam current was measured for a  $90 \times 90 \mu\text{m}$  aperture and for the present aperture. The ratio found was 261 compared to the area ratio of 245. Thus the beam size at the sample is essentially  $6.0 \times 5.5 \mu\text{m}$ . When the pressure vessel was taken to the Cornell High Energy Synchrotron Source, the sample hole position (and shape) was carefully located by studying the change in X-ray transmission as the sample cell was moved in the  $y$  and  $z$  directions (perpendicular to the beam) in 0.1  $\mu\text{m}$  increments.

The cell position was adjusted so that the X-ray beam slightly overlapped the tungsten gasket surrounding the sample in order to obtain a diffraction pattern of the tungsten. Different positions around the sample edge were scanned with no meaningful difference in  $a_r$ .

The enormous stress in the tungsten is far above the expected yield stress of tungsten; at the mean normal stress of 351.2 GPa,  $\sigma_o = 27.9$  GPa. This ensures that hydrogen is present in the sample hole. Theory suggests that the maximum shear strength of perfect crystals is  $G/4$  so that  $\sigma_{00} = G/2 = 81.3$  GPa, while experiment finds for perfect whiskers  $\sigma_{00} = 0.104 G = 16.9$  GPa (ref. 36). Even with these yield stresses the previous conclusion holds; with the enormous stresses in the tungsten and the present geometry the hole would close when the hydrogen leaks away. Moreover, the non-hydrostatic correction to  $P_\sigma$  would be a larger positive correction.

As the skin depth of solid metallic hydrogen at 342 GPa is expected to be  $<0.02 \mu\text{m}$  (based on the resistivity of liquid hydrogen<sup>37</sup>, surely an upper bound) it is clear that no measurable portion of the visible light would pass through the sample layer whose thickness is a few micrometres. We therefore conclude with confidence that, at 300 K and at a pressure of at least 342 GPa, solid hydrogen is not yet an alkali metal.  $\square$



**Figure 3** The effect of the ratio of the initial sample-hole diameter to the flat tip diameter. **a**, Light transmission through the sample at 327 GPa (right). Absorption by the diamonds accounts for the yellow colour. The solid white area (left) shows schematically the initial sample area of 11  $\mu\text{m}$  diameter and the white circle shows the 20- $\mu\text{m}$  tip diameter. The transmission at 342 GPa was similar, as was the transmission after the torque was increased to raise the pressure above 342 GPa (possibly to  $\sim 357$  GPa). Fracture then occurred before the pressure was measured. We emphasize the importance of initially having a small hole diameter relative to the tip diameter. Detailed computer calculations show that if the initial sample hole diameter is 11  $\mu\text{m}$ , the sample area at 342 GPa is 0.30 of the initial sample area as shown schematically in **b**. This compares with the experimental value of 0.29. This leaves 91% of the tip area covered with tungsten. However, if the initial hole diameter equals the tip diameter, the sample diameter first decreases with increasing pressure but then increases; at 173 GPa an instability is reached after which the pressure decreases and the diameter increases on further loading, ending with all the hydrogen beyond the tip and the two diamond tips touching as shown in **c** with none of the tip area covered by tungsten. This catastrophic instability is avoided in the geometry that we used.

Received 31 July 1997; accepted 23 February 1998.

1. Ashcroft, N. W. Pairing instabilities in dense hydrogen. *Phys. Rev. B* **41**, 10963–10971 (1990).
2. Natoli, V., Martin, R. M. & Ceperley, D. M. Crystal structure of atomic hydrogen. *Phys. Rev. Lett.* **70**, 1952–1955 (1993).
3. Loubeyre, P. *et al.* X-ray diffraction and equation of state of hydrogen at megabar pressures. *Nature* **383**, 702–704 (1996).
4. Saumon, D., Chabrier, G., Hubbard, W. B. & Lunine, J. I. in *Strongly Coupled Plasma Physics* (eds Ichimaru, S. & Van Horn, H. M.) 111–120 (Univ. Rochester Press, 1993).
5. Ashcroft, N. W. Metallic hydrogen: a high temperature superconductor. *Phys. Rev. Lett.* **21**, 1748–1749 (1968).
6. Hemley, R. J., Mao, H.-K., Goncharov, A. F., Hanfland, M. & Struzkin, V. Synchrotron infrared spectroscopy to 0.15 eV of  $\text{H}_2$  and  $\text{D}_2$  at megabar pressure. *Phys. Rev. Lett.* **76**, 1667–1670 (1996).
7. Chen, N. H., Sterer, E. & Silvera, I. F. Extended infrared studies of high pressure hydrogen. *Phys. Rev. Lett.* **76**, 1663–1666 (1996).
8. Weir, S. T., Vohra, Y. K. & Ruoff, A. L. Pressure induced metallization of BaSe. *Phys. Rev. B* **35**, 874–876 (1987).

9. Reichlin, R. *et al.* Evidence for the insulator-metal transition in xenon from optical, x-ray and band structure studies to 170 GPa. *Phys. Rev. Lett.* **62**, 669–672 (1989).
10. Goettel, K. A., Eggert, J. H. & Silvera, I. F. Optical evidence for the metallization of xenon at 132 (5) GPa. *Phys. Rev. Lett.* **62**, 665–672 (1989).
11. Luo, H., Desgreniers, S., Vohra, Y. K. & Ruoff, A. L. High pressure optical studies on sulfur to 121 GPa: optical evidence for metallization. *Phys. Rev. Lett.* **67**, 2998–3001 (1991).
12. Desgreniers, S., Vohra, Y. K. & Ruoff, A. L. Optical response of very high density solid oxygen to 132 GPa. *J. Phys. Chem.* **94**, 1117–1122 (1990).
13. Jayaraman, A. Diamond anvil cell and high pressure physical investigations. *Rev. Mod. Phys.* **55**, 65–108 (1983).
14. Kawai, N., Togaya, M. & Mishima, O. A study of the metallic hydrogen. *Proc. Japan Acad.* **51**, 630–633 (1975).
15. Vereshchagin, L. F., Yakovlev, E. N. & Timofeev, Yu A. Possibility of transition of hydrogen into the metallic state. *JETP Lett.* **21**, 85–86 (1975).
16. Hawke, P. S. *et al.* Observation of electrical conductivity of isotropically compressed hydrogen at megabar pressures. *Phys. Rev. Lett.* **4**, 994–997 (1978).
17. Burgess, T. G. & Hawke, R. S. *Metallic Hydrogen Research* 1–26 (Rep. UCID-17977, Lawrence Livermore Lab., CA, 1978).
18. Mao, H. K. & Hemley, R. J. Optical studies of hydrogen above 200 gigapascals: evidence for metallization by band overlap. *Science* **244**, 1462–1465 (1989).
19. Mao, H. K., Hemley, R. J. & Hanfland, N. Infrared reflectance measurements of the insulator-metal transition in solid hydrogen. *Phys. Rev. Lett.* **65**, 484–487 (1990).
20. Hanfland, M., Hemley, R. J. & Mao, H. K. Optical absorption measurements of hydrogen at megabar pressures. *Phys. Rev. B* **43**, 8767–8770 (1991).
21. Eggert, J. H., Goettel, K. A. & Silvera, I. F. High-pressure dielectric catastrophe and the possibility that the hydrogen A-phase is metallic. *Europhys. Lett.* **11**, 775–781 (1980).
22. Ashcroft, N. W. Dense hydrogen: the reluctant alkali. *Phys. World* **8**(7), 43–47 (1995).
23. Weir, S. T., Mitchell, A. C. & Nellis, W. J. Metallization and electrical conductivity of hydrogen in Jupiter. *Phys. Rev. Lett.* **76**, 1860–1863 (1996).
24. Ruoff, A. L., Xia, H., Luo, H. & Vohra, Y. K. Miniaturization techniques for obtaining static pressures comparable to the pressure at the center of the earth: x-ray diffraction at 416 GPa. *Rev. Sci. Instrum.* **61**, 3830–3833 (1990).
25. Ruoff, A. L., Xia, H. & Xia, Q. The effect of a tapered aperture on x-ray diffraction from a sample with a pressure gradient: studies on three samples with a maximum pressure of 560 GPa. *Rev. Sci. Instrum.* **63**, 4342–4348 (1992).
26. Ruoff, A. L. in *High Pressure Science and Technology (Proc. AIRAPT XV and EHRPG 33 Conf.)* (ed. Trzeciakowski, W.) 511 (World Scientific, Warsaw, 1996).
27. Birch, J. F. Finite strain isotherm and velocities for single crystal and polycrystalline NaCl at high pressures and 300 °K. *J. Geophys. Res.* **83**, 1257–1269 (1978).
28. McQueen, R. G., Marsh, S. P., Taylor, J. W., Fritz, J. N. & Carter, W. J. in *High Velocity Impact Phenomena* (ed. Kinslow, R.) 293–417 (Academic, New York, 1970).
29. Carter, W. J., Marsh, S. P., Fritz, J. N. & McQueen, R. J. *Accurate Characterization of the High-Pressure Environment* (ed. Lloyd, E. C.) 147–158 (Spec. Publ. 326, Natl. Bureau of Standards, Washington DC, 1971).
30. Hixson, R. S., Boneness, D. A. & Shaner, J. W. Acoustic velocities and phase transitions in molybdenum under strong shock compression. *Phys. Rev. Lett.* **62**, 637–640 (1989).
31. Christensen, N. E., Ruoff, A. L. & Rodriguez, C. O. Pressure strengthening: a way to multimegabar pressures. *Phys. Rev. B* **52**, 9121–9124 (1995).
32. Ruoff, A. L., Luo, H., Xia, H. & Vohra, Y. K. The effect of nonhydrostaticity on measuring the pressures in metals by energy dispersive x-ray diffraction. *High Press. Res.* **6**, 183–186 (1991).
33. Ruoff, A. L. Penultimate static pressure containment consideration and possible applications to metallic hydrogen preparation. *Adv. Cryogen. Eng.* **18**, 435–440 (1973).
34. Hemley, R. J. *et al.* X-ray imaging of stress and strain of diamond, iron, and tungsten at megabar pressures. *Science* **276**, 1242–1245 (1997).
35. Ruoff, A. L., Luo, H. & Vohra, Y. K. The closing diamond anvil optical window in multimegabar research. *J. Appl. Phys.* **69**, 6413–6416 (1991).
36. Kelly, A. *Strong Solids* 26 (Oxford Univ. Press, New York, 1996).
37. Stevenson, D. J. & Ashcroft, N. W. Conduction in fully ionized liquid metals. *Phys. Rev. A* **9**, 782–789 (1974).

**Acknowledgements.** We acknowledge discussions with N. W. Ashcroft, B. Baranowski and R. M. Martin and thank the CHESS staff for their help. We also acknowledge the role played by B. W. Batterman in creating CHESS. This work was supported by the US NSF and, several years ago in the early stages of this project, by the Cornell Materials Science Center via support by the US NSF.

Correspondence and requests for materials should be addressed to A.L.R. (e-mail: ruoff@msc.cornell.edu).

## Room-temperature transistor based on a single carbon nanotube

Sander J. Tans, Alwin R. M. Verschueren & Cees Dekker

Department of Applied Physics and DIMES, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

The use of individual molecules as functional electronic devices was first proposed in the 1970s (ref. 1). Since then, molecular electronics<sup>2,3</sup> has attracted much interest, particularly because it could lead to conceptually new miniaturization strategies in the electronics and computer industry. The realization of single-molecule devices has remained challenging, largely owing to

difficulties in achieving electrical contact to individual molecules. Recent advances in nanotechnology, however, have resulted in electrical measurements on single molecules<sup>4–7</sup>. Here we report the fabrication of a field-effect transistor—a three-terminal switching device—that consists of one semiconducting<sup>8–10</sup> single-wall carbon nanotube<sup>11,12</sup> connected to two metal electrodes. By applying a voltage to a gate electrode, the nanotube can be switched from a conducting to an insulating state. We have previously reported<sup>5</sup> similar behaviour for a metallic single-wall carbon nanotube operated at extremely low temperatures. The present device, in contrast, operates at room temperature, thereby meeting an important requirement for potential practical applications. Electrical measurements on the nanotube transistor indicate that its operation characteristics can be qualitatively described by the semiclassical band-bending models currently used for traditional semiconductor devices. The fabrication of the three-terminal switching device at the level of a single molecule represents an important step towards molecular electronics.

The samples presented in this study are fabricated as described elsewhere<sup>5,13</sup>. Figure 1a shows an atomic force microscopy (AFM) image of a single nanotube contacting three Pt electrodes (1, 2 and 3). The semiconducting Si substrate, covered with a 300-nm layer of thermally grown SiO<sub>2</sub>, was used as a back-gate (Fig. 1b). In our electrical transport studies on single-wall nanotubes, we have measured many individual tubes (in excess of 20) and find two types of behaviour at room temperature. The metallic variety of tubes reported previously have linear current-voltage curves ( $I-V_{\text{bias}}$ ) and show no dependence on the gate voltage ( $V_{\text{gate}}$ ). Here we present two-probe and three-probe measurements on the second type of sample.

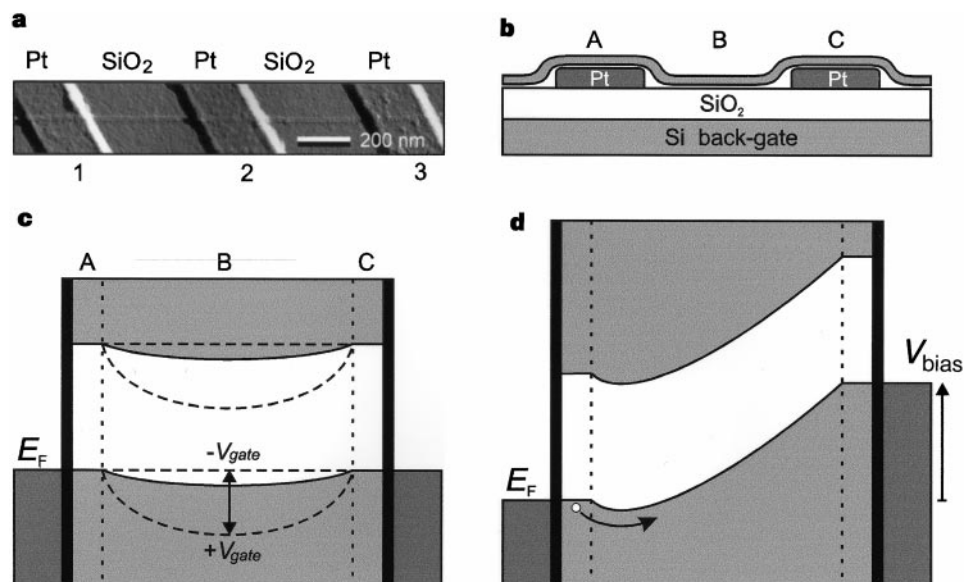
Figure 2 displays  $I-V_{\text{bias}}$  curves for the sample in Fig. 1a. At  $V_{\text{gate}} = 0$ , a small nonlinearity seems to be present in the  $I-V_{\text{bias}}$  curve. When  $V_{\text{gate}}$  is increased to positive values, a pronounced gap-like nonlinearity develops around  $V_{\text{bias}} = 0$ . The curves seem to exhibit a power-law behaviour (solid lines), that is  $I \propto (V_{\text{bias}})^{\alpha}$ , with  $\alpha$  between 1 and 12. Upon application of a negative  $V_{\text{gate}}$ , the  $I-V_{\text{bias}}$  curve becomes linear with a resistance that saturates around 1 M $\Omega$ . This is the same resistance that we find for the metallic tubes in a similar layout. For the major part this resistance is due to the contact resistance between the tube and the electrodes<sup>14</sup>. We thus obtain a controllable semiconductor-to-metal transition in a one-dimensional system. The nonlinearity at room temperature and the asymmetric dependence of the conductance on the gate voltage polarity indicate that the nanotube of this sample is semiconducting. In the inset of Fig. 2, the conductance of the device at  $V_{\text{bias}} = 0$  is plotted against  $V_{\text{gate}}$ . This shows that the conductance can be strongly modulated, by about six orders of magnitude on a change of 10 V in  $V_{\text{gate}}$ . Measurements on the electrode pair 2–3 gave similar results. This sample (Fig. 1a) is one of five samples that showed a similar behaviour. In the other samples, noise due to two-level fluctuators was often larger, especially at large applied voltages. For some samples, the  $I-V_{\text{bias}}$  curves were asymmetric around  $V_{\text{bias}} = 0$ . Although some drift occurred along the  $V_{\text{gate}}$  axis, the transport characteristics of the sample presented here were reproducible over a period of months.

In the energy diagrams of Fig. 1c and 1d we attempt to model the electronic structure and functioning of the TUBEFET (single carbon nanotube field-effect transistor) device. The charge carriers flow through the part of the tube that is on top of the source (A), on the SiO<sub>2</sub> surface (B), and on the drain electrode (C) (see also Fig. 1b). Semiconducting tubes with a 1.4 nm diameter have a bandgap of  $\sim 0.6$  eV (refs 15–17). As for a traditional semiconductor-metal interface, a difference in work function will result in bending of the bands of the semiconductor<sup>18</sup>. The work function of Pt is 5.7 eV, whereas the work function of carbon nanotubes is near 4.5 eV. Owing to this difference, a local polarization layer will develop on



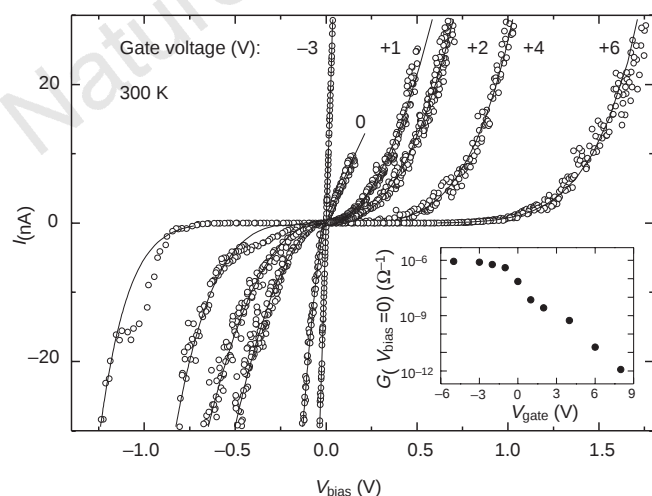
the electrode–nanotube interface until the nanotube valence band edge aligns with the Fermi level of the metal electrode. Such pinning of the valence band edge to the Fermi level of the electrode was observed in scanning tunnelling spectroscopy experiments of semi-conducting nanotubes on Au(111) (ref. 16). Away from the electrodes, the bands bend to lower energy (Fig. 1c, segment B). A gate voltage will not have a strong effect on the nanotube at position A and C

owing to the screening of the nearby metallic leads and the capacitive coupling between the tube and the leads. In segment B, however, the electric field of the gate electrode will couple to the tube. For negative  $V_{\text{gate}}$  this will lead to an accumulation of holes and an increasing conductance, whereas for a positive  $V_{\text{gate}}$  the holes are depleted, yielding a lower conductance (Fig. 1c). Cooling the sample to 160 K, the metallic saturation resistance ( $V_{\text{gate}} < -3$  V)

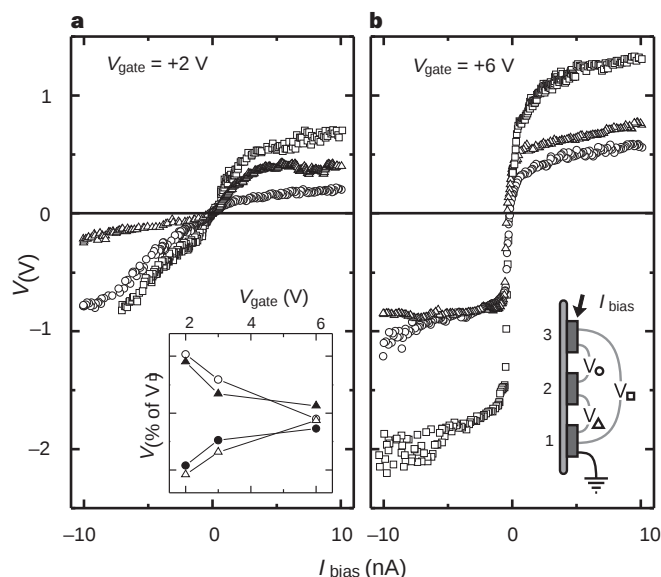


**Figure 1** **a**, Tapping-mode AFM image of an individual carbon nanotube on top of three Pt electrodes. **b**, Schematic side view of the TUBEFET device. A single semiconducting nanotube is contacted by two electrodes. The Si substrate, which is covered by a layer of SiO<sub>2</sub> 300 nm thick, acts as a back-gate. **c**, Suggested band diagram of the device. The nanotube with a gap of ~0.6 eV is connected to the leads with Fermi energy  $E_F$  by tunnelling contacts, indicated by the black

vertical bars. At A and C (see **b**), the valence-band edge is pinned to the Fermi energy of the leads. Owing to a difference in work function between the tube and the electrodes, the bands bend towards lower energy in between the electrodes (B). For positive  $V_{\text{gate}}$  the bands bend more strongly, leading to an insulating state. For negative  $V_{\text{gate}}$  the bands flatten, resulting in a metal-like conductance. **d**, Application of a bias voltage results in a suppression of the barrier.



**Figure 2** Two probe  $I$ - $V_{\text{bias}}$  curves for various values of the gate voltage ( $V_{\text{gate}}$ ). Data were taken at room temperature and in vacuum ( $\sim 10^{-4}$  mbar), with the voltage applied to contacts 1 (drain) and 2 (source). A negative  $V_{\text{gate}}$  leads to ohmic behaviour while a positive  $V_{\text{gate}}$  results in a strong suppression of the current at low bias voltage and nonlinear  $I$ - $V_{\text{bias}}$  curves at higher bias. Inset, conductance at  $V_{\text{bias}} = 0$  as a function of  $V_{\text{gate}}$ . The conductance through this single molecular switch can be varied over at least six orders of magnitude and saturates at  $10^{-6} \Omega^{-1}$  for negative  $V_{\text{gate}}$ .



**Figure 3** Three-probe  $I_{\text{bias}}$ - $V$  curves for two gate voltages. The current is biased over the two outer contacts 1 and 3. The voltage drop is measured over contact pairs 1-2 ( $\Delta$ ), 2-3 ( $\circ$ ) and 1-3 ( $\square$ ) (right inset). At low  $V_{\text{gate}}$  (**a**), the voltage drop over 1-2 is not the same as for 2-3, whereas for larger  $V_{\text{gate}}$  (**b**), the two voltages are identical. Left inset, voltages over pairs 1-2 and 2-3 as a percentage of the total voltage (1-3), both for negative bias (open symbols) and for positive bias (solid symbols). The observed behaviour is in agreement with the proposed band diagram of the device (Fig. 1).

increases from  $\sim 1$  to  $\sim 4$  M $\Omega$ . For a metallic tube we find the same room temperature resistance of  $\sim 1$  M $\Omega$  and a similar dependence on temperature. This result supports the band diagram proposed above (Fig. 1c) because our model predicts that in segments A and C the tube should be metal-like owing to the pinning of the valence band edge. If, as an alternative model, there were an induced Schottky barrier<sup>18</sup> inside the tube at A and C, the device would exhibit a much stronger temperature dependence.

The suggested band structure of this device (Fig. 1c) is similar to that of a traditional semiconductor device, the so-called 'barrier injection transit time' (BARITT) diode<sup>18</sup>. This device consists of a semiconductor connected to two metal contacts, that is, two Schottky-type diodes connected back to back. Although we do not yet have a detailed understanding of the functioning of the TUBEFET, we shall attempt to give a qualitative description by using the well-known BARITT model. In both devices holes have to be transported over the barrier induced by the band bending (in segment B; Fig. 1c). When a bias is applied to a BARITT diode (Fig. 1d) an asymmetric space charge distribution results, because the barrier at the positive contact prevents holes from entering, whereas at the negative contact holes exit from the semiconductor, leaving a large space charge concentration. This leads to an asymmetry in electric field profile and correspondingly yields an unequal voltage drop over the right and left part of the device. When the electric field reaches through the entire device, current is able to flow.

We have performed current-bias transport measurements on the sample of Fig. 1a in a three-point configuration. We now can divide the sample into two segments: one between electrode pair 1–2 and one between pair 2–3. With the current source connected to electrode 3 and electrode 1 grounded, we can measure separately the voltage drops over these two tube segments. The total voltage over pair 1–3 is also measured, with electrode 2 left floating. The result for  $V_{\text{gate}} = 2$  V is displayed in Fig. 3a. Again we observe gap-like features in the  $I$ – $V$  curve. The voltages over the two segments indeed add up to the total voltage drop over the device. However, the voltage drop seems to be different for both segments, with most of the voltage dropping over segment 2–3 at negative bias. This is in agreement with the above described mechanism for a BARITT-like diode because at negative bias the potential of electrode 3 is lower than that of electrode 1, resulting in more space charge and thus a larger voltage drop near contact 3. For positive bias, electrode 1 has the lowest potential, yielding a larger voltage drop at electrode 1. This is indeed observed in the measurements. A different voltage division is observed when the same measurement is performed at a larger positive  $V_{\text{gate}}$  (Fig. 3b). The space-charge profile is now rather symmetric, as can be concluded from the almost equal voltage drop over segments 1–2 and 2–3. At large positive  $V_{\text{gate}}$ , more holes are depleted from the tube, leaving less room for an asymmetric space charge profile. With fewer free charges in the tube, it behaves more like an insulator and the dominant charge build-up is in the electrodes, resulting in equal voltage drops over the two segments. This trend is indicated in the inset of Fig. 3a, where the voltage over the two segments at  $-4$  nA and  $+4$  nA is plotted as a percentage of the total voltage (1–3) against  $V_{\text{gate}}$ .

Screening in truly one-dimensional conductors is expected to be different from that in a three-dimensional system. It is therefore of interest to deduce the typical length over which the bands are bent along the nanotube axis. From our data we can make a rough estimate of this value. Fig. 2 shows that at  $V_{\text{gate}} = 0$  the tube is almost in the metallic state, and that the Fermi energy must therefore be close to the valence band edge throughout the device. If the band-bending length were short ( $\ll 400$  nm), the Fermi level in the nanotube segments away from the electrodes would be located in the middle of the gap between conduction and valence band, and hole transport would be very hard because of the appreciable ( $\sim 0.3$  eV) barrier. If the band-bending length were

very large ( $\gg 400$  nm), however, no significant band bending would occur and the tube would respond similarly to a metallic tube. A negative gate voltage would thus have no effect. What we observe experimentally is something in between, and we thus conclude that, without applied electric fields, the band-bending length is roughly of the order of the distance between the electrodes, which is 400 nm.

The gain of the TUBEFET device can be estimated by considering the voltage over the device at a certain current. Going from  $+4$  V to  $+6$  V in  $V_{\text{gate}}$ , for example, results in a maximum change in voltage over the device of  $\sim 0.7$  V (Fig. 2), which yields a gain of 0.35. The simplest way to improve this value in the current device geometry would be to decrease the SiO<sub>2</sub> layer thickness. Currently this is 300 nm but it could be reduced to  $\sim 5$  nm (ref. 19), resulting in a gain of order 1 or higher. Other geometries that further increase the capacitance between the gate and the tube might also be possible. Considering the ultimate speed of the device, we calculate a ballistic traversal time of  $\sim 10^{-13}$  s (or 10 THz) for 100 nm length of tube and a Fermi velocity for electrons in carbon nanotubes of  $8 \times 10^5$  m s<sup>-1</sup> (ref. 15). Another limitation to the device speed is the RC (resistance–capacitance) time. From Coulomb blockade measurements at low temperatures<sup>5</sup> we deduce that the total capacitance of a piece of nanotube 100 nm long in a similar device geometry is of order  $10^{-18}$  F. With an 'on' resistance of 1 M $\Omega$ , this gives a frequency limit of 0.1 THz. At present the device resistance is dominated by the contact resistance<sup>14</sup>. If low-ohmic contacts can be realized, the two-probe resistance is expected to reach the lower quantum limit of about 6 k $\Omega$ , which would allow a maximum frequency of order 10 THz.

The fabrication of the TUBEFET, the single-molecule field-effect transistor described here, is relatively straightforward, and integration of multiple devices into a circuit may eventually be possible by using molecular self-assembly techniques<sup>20</sup>. Potential applications may therefore be possible, particularly as the device operates at room temperature and high switching speeds, and improved voltage gains seem possible.

In discussions of the fundamental limits of integrated circuit dimensions, warnings are often expressed that at some point radically new device structures should be used because of dominant quantum mechanical effects. Such new device concepts have been explored in the fields of mesoscopic physics and in molecular electronics. It is quite striking that it seems possible to describe qualitatively the ultra-small TUBEFET device, which is based on a single nanotube molecule, by the same semiclassical models that are used for devices in today's computer industry. □

Received 6 January; accepted 2 April 1998.

1. Aviram, A. & Ratner, M. A. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277–283 (1974).
2. Carter, F. L., Siatkowski, R. E. & Wohltjen, H. *Molecular Electronics Devices* (North Holland, Amsterdam, 1988).
3. Aviram, A. (ed.) *Molecular Electronics—Science and Technology* (AIP Conf. Proc. Vol. 262, 1992).
4. Joachim, N., Gimzewski, J. K., Schittler, R. R. & Chavy, C. Electronic transparency of a single C<sub>60</sub> molecule. *Phys. Rev. Lett.* **74**, 2102–2105 (1995).
5. Tans, S. J. *et al.* Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).
6. Porath, D. & Millo, O. Single electron tunneling and level spectroscopy of isolated C<sub>60</sub> molecules. *J. Appl. Phys.* **81**, 2241–2244 (1997).
7. Reed, M. A., Zhou, C., Müller, C. J., Burgin, T. P. & Tour, J. M. Conductance of a molecular junction. *Science* **278**, 252–254 (1997).
8. Mintmire, J. W., Dunlap, B. I. & White, C. T. Are fullerene tubules metallic? *Phys. Rev. Lett.* **68**, 631–634 (1992).
9. Hamada, N., Sawada, A. & Oshiyama, A. New one-dimensional conductors: graphitic microtubules. *Phys. Rev. Lett.* **68**, 1579–1581 (1992).
10. Saito, R., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. Electronic structure of chiral graphene tubules. *Appl. Phys. Lett.* **60**, 2204–2206 (1992).
11. Iijima, S. & Ishihashi, T. Single-shell carbon nanotubes of 1-nm diameter. *Nature* **363**, 603–605 (1993).
12. Bethune, D. S. *et al.* Cobalt-catalysed growth of carbon nanotubes with single-atomic-layer walls. *Nature* **363**, 605–607 (1993).
13. Thess, A. *et al.* Crystalline ropes of metallic carbon nanotubes. *Science* **273**, 483–487 (1996).
14. Bezryadin, A., Verschueren, A. R. M., Tans, S. J. & Dekker, C. Multiprobe transport experiments on individual single-wall carbon nanotubes. *Phys. Rev. Lett.* **80**, 4036–4039 (1998).
15. Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes* (Academic, San Diego, 1996).

16. Wildoer, J., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. Electronic structure of atomically resolved carbon nanotubes. *Nature* **391**, 59–62 (1998).
17. Odom, T. W. *et al.* Atomic structure and electronic properties of single-walled carbon nanotubes. *Nature* **391**, 62–64 (1998).
18. Sze, S. M. *Semiconductor Devices—Physics and Technology* (Wiley, Toronto, 1985).
19. Hirose, H., Hiroshima, M., Yasaka, T., Takakura, M. & Miyazaki, S. Ultra-thin gate oxide growth on hydrogen terminated silicon surfaces. *Microelectron. Eng.* **22**, 3–10 (1993).
20. Ullman, A. *An Introduction to Ultrathin Organic Films* (Academic, San Diego, 1991).

**Acknowledgements.** We thank R. E. Smalley and co-workers for the supply of the indispensable single-wall carbon nanotubes; A. Bezryadin, C.J.P.M. Harms and P. Hadley for discussions; A. van den Enden for technical assistance. The work was supported by the Dutch Foundation for Fundamental Research on Matter (FOM).

Correspondence and requests for materials should be addressed to C.D. (dekker@qt.tn.tudelft.nl).

## Organic-functionalized molecular sieves as shape-selective catalysts

Christopher W. Jones, Katsuyuki Tsuji & Mark E. Davis

Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

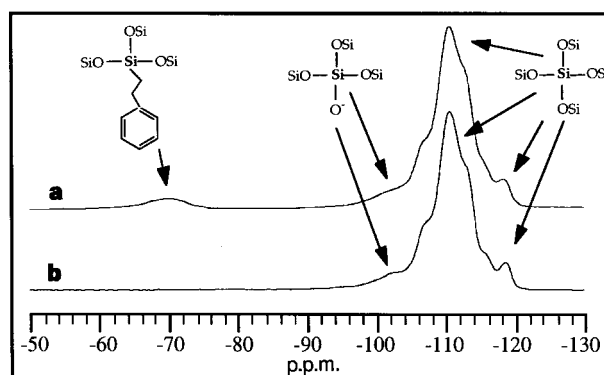
Zeolites and related crystalline molecular sieves can possess catalytically active acid sites, as well as uniformly sized and shaped pores and voids, that allow for their industrial use as shape-selective catalysts<sup>1</sup>. Some catalytic reactions that are not mediated by acids (such as oxidation) have also been shown to occur in zeolites in a shape-selective manner<sup>2</sup>, but the diversity in active sites in these materials remains restricted. For mesoporous materials<sup>3</sup>, the diversity in catalytic activity has been broadened by grafting organosilanes that contain organic functional groups onto the internal pore surfaces<sup>4–6</sup> or by incorporating them into the structure during the synthesis process<sup>7–12</sup>. The former approach has not proven straightforward for microporous zeolites because a large fraction of the grafted functional groups become attached instead to the exterior surfaces of the crystal, where there is no shape selectivity<sup>13</sup>. The synthesis of zeolites and molecular sieves using organosilanes as structure-directing agents has been accomplished<sup>14,15</sup>, but the subsequent creation of porosity requires the complete loss of the organic functional groups. Here we report a new methodology that overcomes these problems and allows the production of microporous molecular sieves containing organic functionalities within their pores. During the initial synthesis phase, phenethyl groups covalently tethered to silicon atoms are incorporated into the framework. The external surface-bound functionalities and the structure-directing agents residing within the intracrystalline spaces are then removed to create a microporous material. Subsequent sulphonation of the phenyl rings produces intrapore sulphonic acid sites that perform shape-selective catalysis. Different active-site types can be created by attaching other functional groups to the framework silicon, and we therefore expect that our method will lead to the formation of a wide range of shape-selective catalysts.

A major key to success in synthesizing organic-functionalized molecular sieves (OFMSs) is the identification of a molecular sieve that can be prepared in the absence of an organic structure-directing agent (SDA) such as NaY, or where the SDA can be removed by extraction. We have synthesized OFMs by both routes, for example NaY (no organic SDA) or pure-silica zeolite beta using tetraethylammonium fluoride (TEAF) as the SDA. We prefer the use of high-silica materials because they provide a hydrophobic void space in which to conduct chemical reactions. An example of the synthetic methodology is illustrated by the synthesis of pure-silica zeolite beta using TEAF as the SDA in the presence of phenethyltrimethoxy silane (PETMS). Following the procedure of Cambor *et al.*<sup>16</sup> for the

synthesis of pure-silica zeolite beta using TEAF as the SDA, we crystallized the hybrid material with 5 atom% or less (2.8 atom% for the sample described here) of the silicon substituted by PETMS. The X-ray diffraction pattern of the hybrid material clearly identifies it as crystalline zeolite beta (diffraction pattern available: see Supplementary information). Figure 1 shows the <sup>29</sup>Si cross-polarization magic angle spinning (CPMAS) NMR spectra for the as-synthesized, pure-silica zeolite beta and the organic-functionalized beta. It is clear from the spectrum of the hybrid material that the phenethyl group (the presence of the aromatic ring was verified by Raman spectroscopy, spectra available, see Supplementary information) is covalently linked to the framework silicon atom (peak at –68 p.p.m.: C–Si–(OSi)<sub>3</sub>; refs 14, 17). This resonance indicates that the linking silicon atom is fully condensed in the framework<sup>18</sup>.

Organic functionalities on the exterior surface of the beta crystals can be removed by reacting the as-synthesized material with concentrated sodium hydroxide solutions (~8 M NaOH, 5% methanol, 25 °C, 1 h). Subsequently, extraction of TEAF from the hybrid material is possible by repeated exposures to acetic acid/water mixtures at 140 °C. Essentially complete removal of TEAF is accomplished for the sample described here (>99% removed as determined by thermogravimetric analysis). The extracted OFMS is sulphonated by contact with vapour from 30% SO<sub>3</sub>/H<sub>2</sub>SO<sub>4</sub> at room temperature after heating at ~100 °C under a vacuum of less than 10<sup>–6</sup> torr overnight. Following sulphonation, the sample is washed with water and then dioxane to remove residual sulphuric acid resulting from the sulphonation procedure. Scanning electron microscopy (SEM) images of this fully modified material do not appear different from the images of the as-made material. Before use as a catalyst, the solid is dehydrated at ~100 °C for at least 6 h under a vacuum of less than 10<sup>–6</sup> torr. Verification of sulphonic acid formation is obtained by Raman spectroscopy (spectrum available, see Supplementary information). Nitrogen adsorption isotherms indicated that the extraction procedure creates significant microporosity (isotherm available, see Supplementary information). Figure 2 illustrates this synthetic procedure.

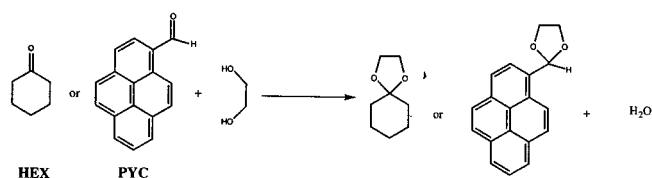
The reaction of a cyclic ketone with ethylene glycol is used to illustrate the catalytic activity and shape-selectivity of the sulphonated, extracted phenyl-functionalized beta (beta/PETMS/SO<sub>3</sub>H). The beta/PETMS/SO<sub>3</sub>H is an active catalyst for the formation of 2,2-pentamethylene-1,3-dioxolane (cyclic ketal) from ethylene glycol and cyclohexanone (HEX). This activity is due to the phenyl-sulphonic acid groups covalently linked to the zeolite framework. The data in Table 1 show that *para*-toluenesulphonic acid monohydrate, phenyl-sulphonic acid anchored to controlled-pore-glass (CPG-240, mean pore diameter 240 Å), and beta/PETMS/SO<sub>3</sub>H are active catalysts for this transformation. None of the OFMS synthesis intermediates and the non-functionalized, pure-silica materials are



**Figure 1** <sup>29</sup>Si CPMAS NMR spectra of **a**, organic-functionalized beta and **b**, pure-silica beta. Spectra are from as-synthesized materials and are referenced to tetramethylsilane.



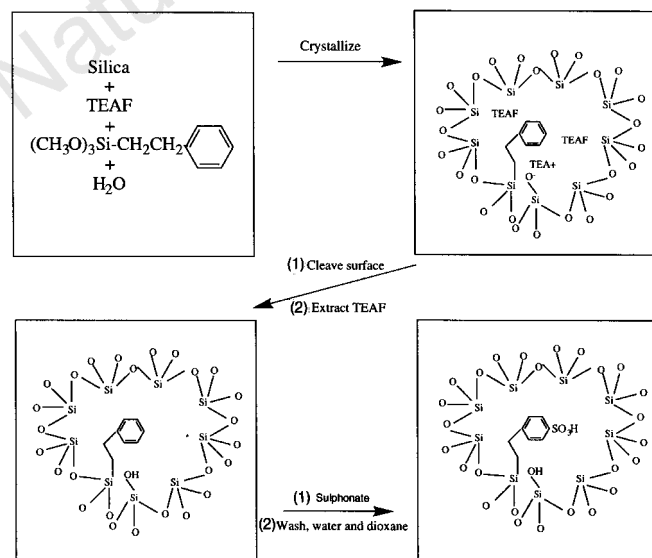
active catalysts. The results clearly indicate that on beta/PETMS/SO<sub>3</sub>H the phenyl-sulphonic acid site is the active centre.



To illustrate shape selectively, 1-pyrenecarboxaldehyde (PYC), which is too large to enter the pore system of beta, is reacted with ethylene glycol. Both HEX and PYC react with ethylene glycol over *para*-toluenesulphonic acid monohydrate to form a ketal or acetal (see Table 1). These two reactions are used to elucidate the location of the phenyl-sulphonic acid moieties in the beta/PETMS/SO<sub>3</sub>H material. Figure 3 illustrates the conversion of PYC or HEX as a function of time using beta/PETMS/SO<sub>3</sub>H as a catalyst. The acetal of PYC is undetectable over the initial 3.5 h of contact. This is due to the lack of sufficient surface catalytic sites that would be required to catalyse the reaction of the bulky PYC within this timeframe. In contrast, HEX is readily converted to its ketal. Upon the addition of di(2-naphthyl)-2-pyrrolidinemethanol (NPM: a bulky poison that cannot access the beta pore system) after 1.15 h of reaction, the conversion of HEX proceeds, thus indicating that the active sites are intrazeolitic. Addition of a small poison that can enter the molecular sieve pores, triethylamine (Et<sub>3</sub>N), stops all reaction at 0.5 h, as indicated by the data in Fig. 3.

As a further control, phenethyl sulphonic acid sites are prepared on the surface of CPG-240. This material has a uniform pore diameter of 240 Å and cannot be a shape-selective catalyst. Over this catalyst, both HEX and PYC are converted, as denoted in Table 1. However, if NPM is added at the outset neither HEX or PYC is reacted. Thus, the shape-selectivity of the OFMS catalyst is demonstrated by the fact that NPM poisons all active sites in CPG-240/PETMS/SO<sub>3</sub>H, but has little effect on beta/PETMS/SO<sub>3</sub>H.

A commercial zeolite beta (PQ, Si/Al = 25) is able to catalyse the



**Figure 2** Schematic illustration of the preparation procedures used to create a sulphonic acid-functionalized molecular sieve.

**Table 1** Ketal or acetal formation with cyclohexanone or 1-pyrenecarboxaldehyde and ethylene glycol

| Catalyst                                      | Conversion (%) of HEX |
|-----------------------------------------------|-----------------------|
| None                                          | <2                    |
| CPG-240                                       | <2                    |
| CPG-240/PETMS                                 | <2                    |
| CPG-240/PETMS/SO <sub>3</sub> H†              | 71.0                  |
| Si-beta, as made                              | <2                    |
| Si-beta, extracted                            | <2                    |
| Si-beta, sulphonated                          | <2                    |
| Si-beta/PETMS, as-made                        | <2                    |
| Si-beta/PETMS, as-made sulphonated            | <2                    |
| Si-beta/PETMS, extracted                      | <2                    |
| Si-beta/PETMS, extracted, SO <sub>3</sub> H†‡ | 72.0                  |
| <i>Para</i> -toluenesulphonic acid†           | 70.0                  |
| <i>Para</i> -toluenesulphonic acid, NPM§      | 0                     |
| <i>Para</i> -toluenesulphonic acid†           | 40.0                  |
| CPG-240/PETMS/SO <sub>3</sub> H, NPM§         | 0                     |
| CPG-240/PETMS/SO <sub>3</sub> H†              | 40.0                  |

\* CPG-240/PETMS/SO<sub>3</sub>H: ~0.13 mmol H<sup>+</sup> g<sup>-1</sup> catalyst.

‡ Si-beta/PETMS, extracted, SO<sub>3</sub>H: ~0.14 mmol H<sup>+</sup> g<sup>-1</sup> catalyst.

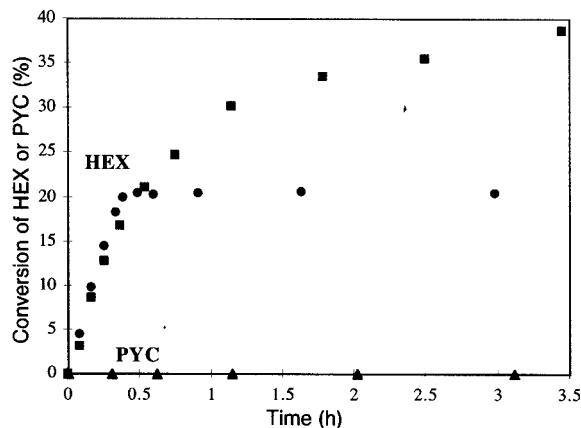
† Conversions are limited by equilibrium.

§ No conversion for both HEX and PYC. NPM poisons all active sites in CPG-240 and *para*-toluenesulphonic acid, unlike beta.

|| Conversion of PYC.

reaction demonstrated here. However, in the absence of the surface poison NPM, there is significant conversion of PYC by this catalyst (initial rates: HEX/PYC = 16, same conditions as those for experiments described in Fig. 3). Thus, the aluminosilicate is not as shape-selective because of the reactivity of the external crystal surface. Also, there are no proven examples of shape-selective catalysis with organic-functionalized mesoporous solids. The pores of the mesoporous MCM-41-type materials can be constricted into the micropore range by silanation treatments. However, these modifications will probably not result in a single, fixed-diameter pore opening as found in crystalline silicate molecular sieves. Instead, a distribution of pore sizes would be expected. In addition, the crystalline OFMSs are physically and chemically more robust than amorphous, mesoporous materials.

The organic-functionalized materials of the type described here provide new opportunities for shape-selective catalysis. In principle, any functional group that catalyses homogeneous reactions can be 'tailor-made' into an intrazeolitic organic moiety. Thus far, we have prepared OFMS using zeolite beta (TEAF as SDA), ZSM-5 (hexamethylenediamine as SDA) and NaY and have placed numerous



**Figure 3** Shape-selective catalysis over organic-functionalized beta. Reactions of PYC (▲) and HEX (■, ●); 27.5 mg NPM added at 1.15 h for the experiment denoted by ■ and 100 mg Et<sub>3</sub>N added at 0.5 h for the experiment denoted by ●. Acid content of catalyst: ~0.14 mmol H<sup>+</sup> g<sup>-1</sup> catalyst.

functional groups into their structures, for example phenethyl, cyclohexenyl, aminopropyl (information regarding zeolite NaY with tethered cyclohexenylethyl groups is included in Supplementary information). Additionally, we have demonstrated not only acid catalysis, but base-catalysed conversions. Finally, the organic moieties may perform functions other than catalysis, for example, synthesis of vicinal diols through oxidation of tethered cyclohexenyl groups could be used as a chelating site for adsorbing metal ions from solution. We believe that this synthetic method should be a general method for high-silica molecular sieve syntheses, and particularly syntheses mediated by  $F^-$ , so long as the SDA is extractable and there exists sufficient intracrystalline space for both the organic functional group and the SDA. Thus, organic-functionalized molecular sieves should provide for new applications of shape-selectivity. □

## Methods

**Catalytic reactions.** For Table 1. Reactions were conducted in magnetically stirred glass reactors at 70 °C for 24 h. The reactor was charged with 9 g toluene, 10 mmol of each reactant, and about 10 mg catalyst. Products were identified by gas chromatography using authentic samples.

For Figure 3. Reactions were conducted in magnetically stirred glass reactors at 70 °C. The reactor was charged with 10 g toluene, 10 mmol of reactants in the case of HEX or 3 mmol in the case of PYC, and 13 mg catalyst. Samples were taken periodically and analysed by GC/MS spectroscopy. Acid content of catalyst:  $\sim 0.14 \text{ mmol H}^+ \text{ g}^{-1}$  catalyst.

**Active site density.** Beta/PETMS/ $\text{SO}_3\text{H}$  (0.21 g) was washed with 15 ml saturated NaCl solution at room temperature. The OFMS was removed by filtration. Several drops of phenolphthalein solution were added to the filtrate and then this solution was titrated with 0.001 M NaOH to neutrality. The active-site density obtained from this method agreed well with the total mass of organic material as determined by TGA, and elemental analysis for sodium.

Received 7 November 1997; accepted 17 March 1998.

1. Venuto, P. N. Organic catalysis over zeolites: a perspective on reaction paths within micropores. *Microporous Mater.* **2**, 297–411 (1994).
2. Tatsumi, T., Yanagisawa, K., Asano, K., Nakamura, M. & Tominaga, H. in *Studies in Surface Science and Catalysis 83* (eds Hattori, T. et al.) 417–424 (Elsevier, Amsterdam, 1994).
3. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
4. Corma, A., Iglesias, M., del Pino, C. & Sanchez, F. New rhodium complexes anchored on modified USY zeolites. *J. Chem. Soc. Chem. Commun.* 1253–1255 (1991).
5. Brunel, D., Cauvel, A., Fajula, F. & DiRenzo, F. in *Studies in Surface Science and Catalysis 97* (eds Bonnevot, L. et al.) 173–180 (Elsevier, Amsterdam, 1995).
6. Diaz, J. J., Balkus, K. J. Jr, Bedioui, F., Kurshev, V. & Kevan, L. Synthesis and characterization of cobalt-complex functionalized MCM-41. *Chem. Mater.* **9**, 61–67 (1997).
7. Burkett, S. L., Sims, S. D. & Mann, S. Synthesis of hybrid inorganic-organic mesoporous silica by co-condensation of siloxane and organosiloxane precursors. *Chem. Commun.* 1367–1368 (1996).
8. Lim, M. H., Blanford, C. F. & Stein, A. Synthesis and characterization of a reactive vinyl-functionalized MCM-41: probing the internal pore structure by a bromination reaction. *J. Am. Chem. Soc.* **119**, 4090–4091 (1997).
9. Fowles, C. E., Burkett, S. L. & Mann, S. Synthesis and characterization of ordered organo-silica-surfactant mesophases with functionalized MCM-41-type architecture. *Chem. Commun.* 1769–1770 (1997).
10. Macquarrie, D. J. Direct preparation of organically modified MCM-type materials: preparation and characterization of aminopropyl-MCM and 2-cyanoethyl-MCM. *Chem. Commun.* 1961–1962 (1996).
11. Lim, M. H., Blanford, C. F. & Stein, A. Synthesis of ordered microporous silicates with organo sulfur surface groups and their applications as solid acid catalysts. *Chem. Mater.* **10**, 467–470 (1998).
12. Itoh, A. & Masaki, Y. Synthesis and application of new phenyl-functionalized zeolites as protection against radical bromination at the benzylic position. *Syn. Lett.* **12**, 1450–1452 (1997).
13. Cauvel, A., Brunel, D., DiRenzo, F., Moreau, P. & Fajula, F. in *Studies in Surface Science and Catalysis 94* (eds Beyer, H. K. et al.) 286–293 (Elsevier, Amsterdam, 1995).
14. Li, H. X., Cambor, M. A. & Davis, M. E. Synthesis of zeolites using organosilicon compounds as structure-directing agents. *Microporous Mater.* **3**, 117–121 (1994).
15. Calabro, D. C. *US Patent 5,194,410* (1993).
16. Cambor, M. A., Corma, A. & Valencia, S. Spontaneous nucleation and growth of pure silica zeolite beta free of connectivity defects. *Chem. Commun.* 2365–2366 (1996).
17. Sudholter, E. J. R., Huis, R., Hays, G. R. & Alma, N. C. M. Solid state silicon-29 and carbon-13 NMR spectroscopy using cross-polarization and magic-angle spinning techniques to characterize 3-chloropropyl and 3-aminopropyl-modified silica gels. *J. Coll. I. Sci.* **103**, 554–560 (1985).
18. Bein, T., Carver, R. F., Farlee, R. E. & Stucky, G. D. Solid-state  $^{29}\text{Si}$  NMR and infrared studies of the reactions of mono and polyfunctional silanes with zeolite Y surfaces. *J. Am. Chem. Soc.* **100**, 4546–4553 (1978).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Correspondence and requests for materials should be addressed to M.E.D. (e-mail: mdavis@chem.caltech.edu).

## Correlation between Arabian Sea and Greenland climate oscillations of the past 110,000 years

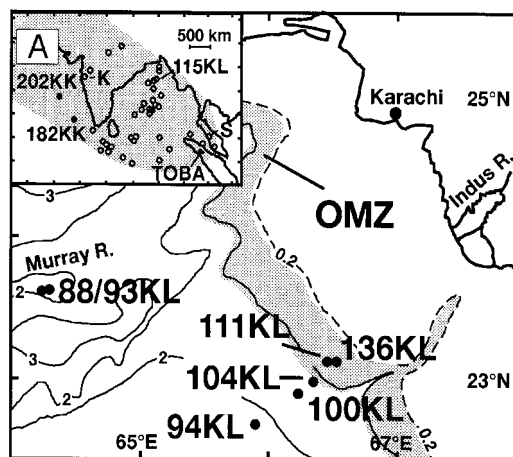
Hartmut Schulz\*, Ulrich von Rad\* & Helmut Erlenkeuser†

\* Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), PF 510153, D-30631 Hannover, Germany

† Leibniz-Labor für Altersbestimmung und Isotopenforschung, Universität Kiel, Max-Eyth-Strasse 11-13, D-24118 Kiel, Germany

Palaeoclimate studies have revealed the general high-frequency instability of Late Pleistocene climate—for example, the so-called Dansgaard-Oeschger and Heinrich events—on timescales of a few millennia, centuries or even decades<sup>1–11</sup>. Here we present evidence for a general relationship between low-latitude monsoonal climate variability and the rapid temperature fluctuations of high northern latitudes that are recorded in the Greenland ice records. Sediment cores from the northeastern Arabian Sea show laminated, organic-carbon-rich bands, reflecting strong monsoon-induced biological productivity, that correlate with the mild interstadial climate events in the northern North Atlantic region. In contrast, periods of lowered southwest monsoonal intensity, indicated by bioturbated, organic-carbon-poor bands, are associated with intervals of high-latitude atmospheric cooling and the injection of melt water into the North Atlantic basin. Our records suggest that Dansgaard-Oeschger and Heinrich events are strongly expressed in low-latitude (monsoonal) climate variability, suggesting the importance of common forcing agents such as atmospheric moisture and other greenhouse gases.

Today, the northern Arabian Sea is characterized by (1) warm and highly saline near-surface waters originating from the Persian Gulf,

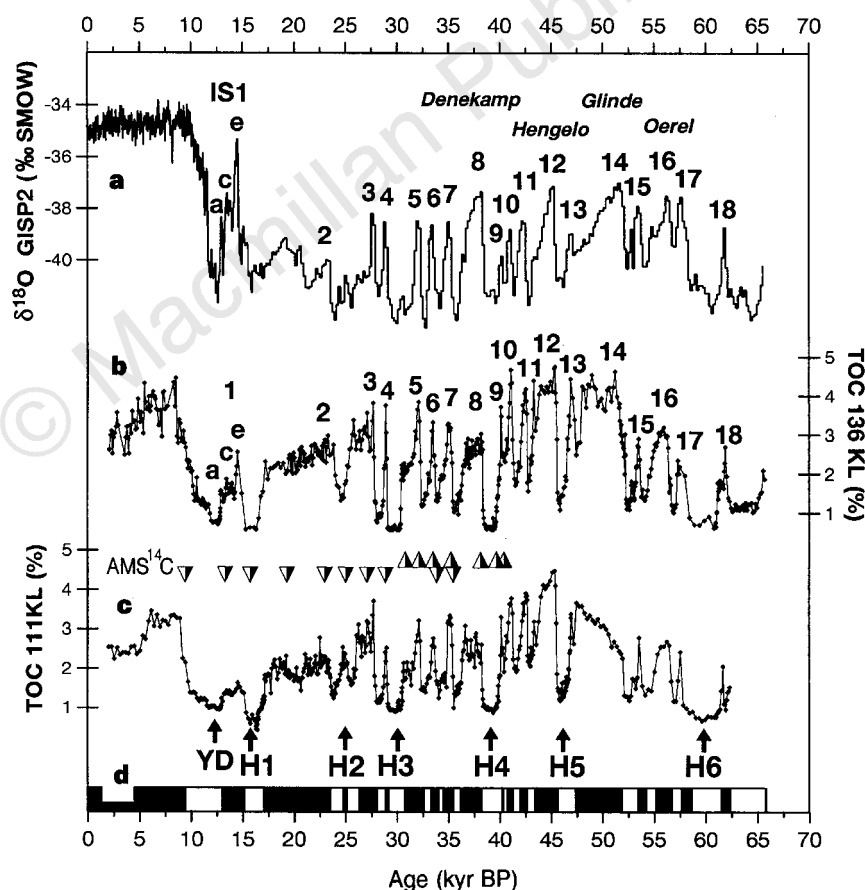


**Figure 1** Location of sediment cores off Pakistan. OMZ, oxygen-minimum zone impinging on the continental slope between 200 and 1,200 m water depth (w.d.) with contours in km. Labelled sites and key cores SO90-88KL and SO90-93KL (23°35' E, 64°13' E, 1,783 m and 1,802 m w.d.), SO90-111KL (23°06' N, 66°29' E, 775 m w.d.), SO90-136KL (23°07' N, 66°30' E, 568 m w.d.) are discussed in the text. Inset map showing sites of youngest Toba ash occurrences between Toba (Sumatra) and the northeastern Arabian Sea. 'A' indicates position of main map. Open circles, localities from the Bay of Bengal, India (K indicates Kukdi) and southern Malaysia (S indicates Serdang)<sup>23–25</sup> (H. R. Kudrass, personal communication); filled circles: Arabian Sea IIOE cores<sup>12</sup>.

(2) high fluxes of both wind- and river-transported (Indus) terrestrial sediment, and (3) high oceanic (monsoonal) surface-water productivity and biogenic fluxes, leading to (4) a stable, mid-depth oxygen-minimum zone (OMZ) at 200–1200 m water depth (Fig. 1). The OMZ prevents bioturbation and favours the accumulation of organic-carbon-rich, partly laminated or even varved sediments<sup>12</sup> (here we will use TOC to denote total organic carbon). These conditions offer a unique opportunity to study in detail the variability of the tropical monsoonal circulation and marine productivity on millennial to annual timescales during the last glacial–interglacial cycle<sup>13</sup>. In contrast to the northwestern Arabian Sea where a seasonal reversal in surface circulation occurs due to the strong summer southwestern-monsoon winds (July–August) and (less intense) northeastern winds during winter (December–February), the seasonal variation of current directions is less distinct off Pakistan. Although geochemical analyses show that the organic matter fraction is predominantly of marine origin, the stable oxygen isotopes of planktonic foraminifera in the surface sediments do not show a clear upwelling signal of cool subsurface waters. We infer that the seasonal productivity in the northeastern Arabian Sea is temporally and spatially linked to the strong coastal and open-ocean

upwelling cells off Oman by the lateral advection of nutrient-rich waters from the southwest due to the anticyclonic surface gyre during the summer months<sup>14</sup>.

The sediment cores taken from OMZ water depths along the continental margin off Pakistan (Fig. 1) reveal a consistent pattern of dark, TOC-rich, partly laminated intervals alternating with light-coloured, TOC-poor, pteropod-rich, bioturbated intervals (Fig. 2). TOC concentrations in the hemipelagic sediments of cores SO90–88/93KL from the Murray ridge below the present OMZ are generally much lower, and show significant variability only on the longer-term, orbitally induced timescales. However, these cores also document nearly identical high-frequency variability, seen in distinct peaks of pteropod debris and increased biogenic sand content corresponding to the bioturbated, pteropod-rich and TOC-poor intervals in the OMZ cores. Here we use the record of sonic velocity ( $V_p$ ) reflecting the bulk compositional changes in the sediment. Based on geochemical and micropalaeontological investigations and on the spatial distribution of laminated sediments from box core tops from this area<sup>13</sup>, we suggest that the multiple facies changes in our cores reflect fluctuations in the intensity of summer monsoonal productivity. Apparently, the oxygenation of



**Figure 2** Correlation of high-frequency climate variability for the past 65,000 years between ice-core and marine-sediment-core records. **a**, Greenland GISP2  $\delta^{18}\text{O}_{\text{ice}}$  record<sup>1</sup>; **b** and **c**, two high-resolution (average sample resolution  $\sim 100$  yr) marine TOC-records of cores SO90–111KL and SO90–136KL off Pakistan, based on a 1:1 fit using the GISP2 chronology. Following the generalized lithology shown in **d**, profiles from the centre of the OMZ are characterized by conspicuous alternations between dark-coloured, distinctly to indistinctly laminated, TOC-rich (black in **d**) and light-coloured, bioturbated pteropod-rich and TOC-poor intervals (white in **d**), indicating millennial- to centennial-scale variability of monsoonal surface water productivity and bottom-water oxygenation. Numbers indicate Greenland interstadials IS1–18, and equivalent Arabian Sea monsoonal events 1–18; H1–H6 indicate northern North Atlantic Heinrich meltwater events, coinciding with the

deposition of bioturbated intervals in the Arabian Sea record. YD, Younger Dryas; triangles in **b** and **c** show AMS  $^{14}\text{C}$  dates. Total organic carbon (TOC, as percentage of dry weight) contents were calculated from sediment colour (grey-scale) measurements. Based on 138 and 136 paired samples, polynomial regression equations revealed  $r^2 = 0.93$  and  $s = 0.35\%$  for cores 111KL and 136KL/137KA<sup>13</sup>, respectively. For AMS dating, monospecific samples of planktic foraminiferal species were mostly taken from the laminated sections where biases due to bioturbational mixing can be ruled out. Radiocarbon ages were calibrated to a calendar timescale<sup>29</sup> (except for the youngest sample<sup>30</sup>), after a correction of  $\sim 640$  yr for the local marine reservoir effect determined from varved surface sediments. These data are available; see Supplementary information.

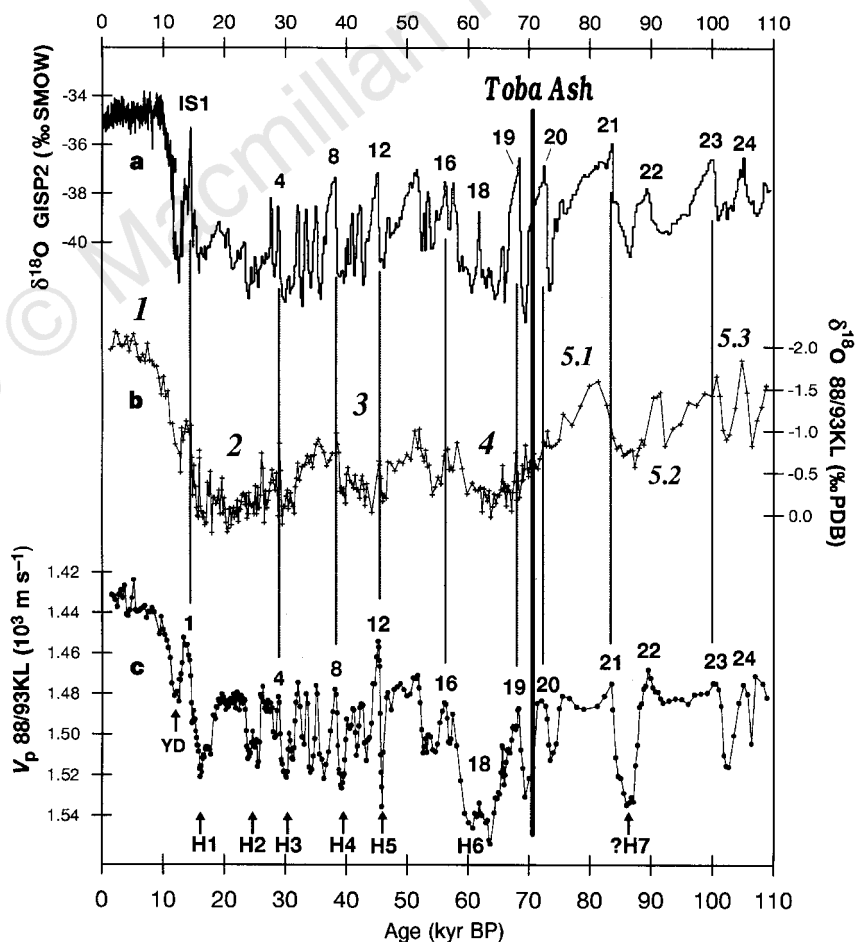


bottom waters is directly influenced by these fluctuations resulting in strong variations of the intensity of the OMZ during the last glacial period.

Our records from the OMZ off Pakistan show striking similarities with the record of  $\delta^{18}\text{O}$  from the Greenland GRIP and GISP2 ice cores that documents extremely rapid fluctuations in air temperature in the high northern latitudes during the last glacial–interglacial cycle<sup>1,2</sup>. Generally, the light-coloured, bioturbated intervals correspond to periods of extremely ‘light’ ( $^{18}\text{O}$ -depleted)  $\delta^{18}\text{O}$  values in the GISP2 ice core indicating cool air temperatures over Greenland and the northern North Atlantic Ocean (Fig. 2). This applies particularly for the six intervals of bioturbated,  $\text{CaCO}_3$ - and pteropod-rich sediments in the OMZ cores which can be correlated to the Younger Dryas cool period at around 12,600 years BP and to the Heinrich events H1–H6 of ice-rafted detritus that were associated with massive discharges of melt water into the North Atlantic<sup>3,4</sup>. Conversely, the dark-coloured, indistinctly to distinctly laminated intervals in the Arabian Sea sediments appear as equivalents to all Greenland Dansgaard–Oeschger interstadials IS1 to IS18. Similar to the Greenland ice record, there is a clear pattern of two intervals of prominent triple peaks of high organic carbon content, IS5, IS6, IS7, and IS9, IS10, IS11. Both triple peaks are preceded by broader events corresponding to IS8 (Denekamp) and IS12 (Hengelo). For the early marine oxygen isotope stage 3, stage 4

and the late stage 5 (Figs 2 and 3), bundles of intervals can be grouped as equivalents to events IS13, IS14 (Glinde), IS15, IS16, IS17 (Oerel), to IS21 (Odderade), and to IS22, IS23, IS24 (Brørup) according to the nomenclature of the northwest European interstadial periods<sup>1</sup>. The Atlantic-type structure of the Bølling/Allerød climate interval (IS1, a, c, e) consisting of three ‘cool’ and four ‘warm’ episodes<sup>15</sup> is also recorded.

High-frequency monsoonal variability in the properties of surface waters is also documented in the stable-isotope record of the near-surface-dwelling foraminiferal species *Globigerinoides ruber* (Fig. 3b). The glacial to interglacial  $\delta^{18}\text{O}$  amplitude is  $\sim 2\text{‰}$ . The global ice-volume effects of  $\sim 1.3\text{‰}$  leaves up to  $0.7\text{‰}$  to be explained by the local variability of water temperature of the order of  $\pm 3\text{--}4^\circ\text{C}$  and/or salinity of  $\pm 1\text{--}1.5\text{‰}$ . We observe a strong long-term fluctuation in our record on the 23,000-year precessional cycle of solar radiation, forcing the monsoons<sup>16</sup>, as broad  $\delta^{18}\text{O}$  maxima are centred around 20,000, 42,000, 64,000 and 87,000 yr BP. For instance, the broad mid-stage-3 maximum at 42,000 yr BP is of regional significance and is not seen in the standard marine  $\delta^{18}\text{O}$  record that is used for global chronostratigraphic correlation<sup>17</sup>. Superimposed, further distinct millennial- to centennial-scale  $\delta^{18}\text{O}$  maxima indicate periods of high surface salinity and/or lowered temperature that are associated with the equivalents of the Heinrich events seen in the  $V_p$ -record. Conversely, nearly all



**Figure 3** Correlation of high-frequency climate variability for the past 110,000 yr between ice-core and marine-sediment-core records. **a**, Greenland GISP2  $\delta^{18}\text{O}_{\text{ice}}$  record; **b**, the marine  $\delta^{18}\text{O}$  stable-isotope record of the planktic foraminifer *Globigerinoides ruber*; **c**, the record of sonic velocity ( $V_p$ ) of sediment cores SO90-88/93KL. Notation of interstadials IS1–IS24, YD, Heinrich events H1–H7 and Arabian Sea equivalent events 1–24 follows Fig. 2. Italic numbers 1–5.3 indicate

standard SPECMAP<sup>17</sup> oxygen-isotope stages. In cores 88KL and 93KL, shards derived from the Toba mega-eruption occur at 620–624 and 623–627 cm core depth, respectively, near the isotope stage 4/5 boundary, that is, between Greenland interstadials IS19 and IS20 as well as between Arabian Sea equivalents 19 and 20.

interstadial equivalents seem to be related with minima in the  $\delta^{18}\text{O}$  record. A series of similar century-scale  $\delta^{18}\text{O}$ -events at the glacial/interglacial transition is reported from the northwestern Arabian Sea, suggesting that the onset and intensification of southwest-monsoonal circulation and surface water productivity occurred at discrete steps at the end of the last glaciation<sup>18</sup>, in phase with the onset and subsequent events of deglacial warming in the high latitudes<sup>19</sup>.

Confirmed by evidence from seventeen AMS-radiocarbon ages and high-resolution stable-isotope stratigraphy we suggest that our sediment cores show for the past 110,000 years a record of low-latitude monsoonal variability to be correlated to the cool Younger Dryas and Heinrich events H1–H7, and to the interglacial/interstadial Holocene, Bølling/Allerød and Dansgaard–Oeschger IS1–IS24 climate intervals, respectively, seen in the Greenland temperature record.

Our correlations are further supported by the identification of the youngest Toba ash horizon in the Arabian Sea sediments, also manifested as a distinct aerosol peak in the Greenland GISP2 ice core ~71,000 years ago, between interstadials IS19 and IS20<sup>20</sup>. The Toba eruption, the largest Quaternary volcanic eruption on Earth, produced more than 800 km<sup>3</sup> of rhyolitic bubble-wall shards and pumice which were deposited during a few weeks covering at least 1% of the Earth's surface with an extensive ash blanket<sup>21,22</sup>. Owing to the prevailing tropospheric jet streams, the youngest Toba ash was mainly dispersed to the west-northwest. Up to now particles have been traced from Malaysia to the Bay of Bengal up to western India, ~3,100 km from the source<sup>23–25</sup> (Fig. 1). By careful inspection of our records from the International Indian Ocean Expedition and off Pakistan<sup>12,13,14</sup>, the youngest Toba ash was identified in the Arabian Sea cores IIOE-182KK and –202KK, in SO90-94KL, –100KL and –93KL from the northeastern Murray ridge. These sites extend the maximum distance of the Toba ash occurrences to more than 4,000 km from the source and the area of the ash blanket to  $> 4 \times 10^6 \text{ km}^2$ .

We investigated the composition of the volcanic particles in Arabian Sea core SO90-94KL where a disseminated (bioturbated) ash was found at 520 cm core depth consisting of fresh, transparent bubble-wall shards and in Bengal fan core SO93-115KL (Fig. 1) from a discrete 6-cm-thick ash layer (H. R. Kudrass, personal communication). The chemical composition was determined by analysing discrete glass shards by electron microprobe. The Arabian Sea results show very similar major-element contents to the Bengal fan samples, and correlate well with the composition from the distal Bengal fan, western India, ODP Site 758 on the outermost Bengal fan, as well as the most proximal ash at Malaysia. The chemical results on the youngest Toba ash composition correlated from Sumatra to the Arabian Sea are available (see Supplementary Information).

In cores SO90-88/93KL, rhyolitic shards occur at narrow intervals of a few centimetres near the boundary between marine isotope stages 4 and 5 (Fig. 3), in a similar position to the records from the Bay of Bengal<sup>23</sup>. As cores SO90-88/93KL are situated in an isolated position on the Murray ridge, we can exclude downslope reworking. We suggest that the glass shards were rapidly deposited as fallout ash and then slightly disseminated due to bioturbational mixing. Based on  $\delta^{18}\text{O}$  stratigraphy we estimate ages of ~72,400 to 74,600  $\pm$  5,000 yr BP for the Toba ash in the Arabian Sea cores, in good agreement with the ages centred at 74,000  $\pm$  2,000 yr BP based on K/Ar, <sup>40</sup>Ar/<sup>39</sup>Ar, and fission track dating of glass shards from terrestrial outcrops<sup>23,24</sup>. In contrast, the SO<sub>4</sub><sup>2–</sup> peak attributed to the Toba mega-eruption in the GISP2 core from Greenland has been attributed a slightly younger age of ~71,100  $\pm$  5,000 yr BP, using the correlated age model from the two  $\delta^{18}\text{O}_{\text{ice}}$ -records of GISP2 and Vostok<sup>26</sup>. However, the Toba ash as an isochronous stratigraphic horizon directly correlates the Arabian Sea and Greenland records

far beyond the reach of radiocarbon dating (Fig. 3). In the GISP2-record the Toba signal is placed at the late IS20; in cores SO90-88/93KL the Toba ash occurs between the equivalent monsoonal events 19 and 20.

Our records from the northwestern Indian Ocean suggest that the rapid Dansgaard–Oeschger- and Heinrich-style events known from Greenland and the northern North Atlantic are a significant component also of low-latitude climate variability during the past 110,000 years, due to millennial to centennial fluctuations in the intensity of southwest monsoonal circulation. These rapid fluctuations point to large-scale climate elements in the ocean–atmosphere system interacting rapidly in the high and low latitudes. We suggest that the record of mild interstadials found in the Greenland ice cores reflect periods of global warmth, associated with high levels of tropospheric moisture and other greenhouse gases<sup>27,28</sup>. □

Received 15 May 1997; accepted 2 February 1998.

- Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
- Johnsen, S. J. *et al.* Irregular glacial interstadials recorded in a new Greenland ice core. *Nature* **359**, 311–313 (1992).
- Broecker, W., Bond, G., Klas, M., Clark, E. & McManus, J. Origin of the northern Atlantic Heinrich events. *Clim. Dyn.* **6**, 265–273 (1992).
- Grousset, F. *et al.* Patterns of ice-rafted detritus in the glacial North Atlantic. *Paleoceanography* **8**, 175–192 (1993).
- Kotilainen, A. & Shackleton, N. J. Rapid climate variability in the North Pacific Ocean during the past 95,000 years. *Nature* **377**, 323–326 (1995).
- Tada, R., Irino, T. & Koizumi, I. in *Proc. of the 1994 Sapporo IGBP Symp., Global Fluxes of Carbon and Its Related Substances in the Coastal Sea-Ocean-Atmosphere System* (ed. Tsunogai, S.) 517–522 (Sapporo, 1995).
- Porter, S. C. & An, Z. S. Correlation between climate events in the North Atlantic and China during the last glaciation. *Nature* **375**, 305–308 (1995).
- Behl, R. J. & Kennett, J. P. Brief interstadial events in the Santa Barbara basin, NE Pacific during the past 60 kyr. *Nature* **379**, 243–246 (1996).
- Mayewski, P. A. *et al.* Climate change during the Last Deglaciation in Antarctica. *Science* **272**, 1636–1638 (1996).
- Bender, M. *et al.* Climate correlations between Greenland and Antarctic during the past 100,000 years. *Nature* **372**, 663–666 (1994).
- Jouzel, J. Ice cores north and south. *Nature* **372**, 612–613 (1994).
- von Stackelberg, U. Faziesverteilung in Sedimenten des indisch-pakistanischen Kontinentalrandes (Arabisches Meer). “Meteor”-Forsch.-Ergebn., Reihe C **9**, 1–73 (1972).
- von Rad, U., Schulz, H. & SONNE 90 Scientific Party Sampling the oxygen minimum zone off Pakistan: glacial/interglacial variations of anoxia and productivity. *Mar. Geol.* **125**, 7–19 (1995).
- Schulz, H., von Rad, U. & von Stackelberg, U. in *Paleoclimatology and Paleoceanography from Laminated Sediments* (ed. Kemp, A. E. S.) 185–207 (Geol. Soc. Spec. Publ. No. 116, Blackwell Scientific, Oxford, 1996).
- Hughen, K., Overpeck, J. T., Peterson, L. C. & Trumbore, S. Rapid climate changes in the tropical Atlantic region during the last deglaciation. *Nature* **380**, 51–54 (1996).
- Clemens, S., Prell, W., Murray, D., Shimmield, G. & Weedon, G. Forcing mechanisms of the Indian ocean monsoon. *Nature* **353**, 720–725 (1991).
- Martinson, D. G. *et al.* Age dating and the orbital theory of the ice ages: development of a high-resolution 0 to 300,000 year chronostratigraphy. *Quat. Res.* **27**, 1–29 (1987).
- Sirocko, F. *et al.* Century-scale events in monsoonal climate over the past 24,000 years. *Nature* **364**, 322–324 (1993).
- Sirocko, F., Garbe-Schönberg, D., McIntyre, A. & Molino, B. Teleconnections between the subtropical monsoons and high-latitude climates during the last deglaciation. *Science* **272**, 526–529 (1996).
- Zielinski, G. A. *et al.* Potential atmospheric impact of the Toba mega-eruption 71,000 years ago. *Geophys. Res. Lett.* **23**, 837–840 (1996).
- Rose, W. I. & Chesner, C. A. Dispersal of ash in the great Toba eruption, 75 ka. *Geology* **15**, 913–917 (1987).
- Rose, W. I. & Chesner, C. A. Worldwide dispersal of ash and gases from earth's largest known eruption: Toba, Sumatra, 75 ka. *Paleogeogr. Paleoclimatol. Paleoecol.* **89**, 269–275 (1990).
- Ninkovich, D., Shackleton, N. J., Abdel-Monem, A. A., Obradovich, J. D. & Izett, G. K-Ar age of the Pleistocene eruption of Toba, north Sumatra. *Nature* **276**, 574–577 (1978).
- Chesner, C. A., Rose, W. I., Deino, A., Drake, R. & Westgate, J. A. Eruptive history of Earth's largest Quaternary caldera (Toba, Indonesia) clarified. *Geology* **19**, 200–203 (1991).
- Acharyya, S. K. & Basu, P. K. Toba ash on the Indian subcontinent and its implications for the correlation of late Pleistocene alluvium. *Quat. Res.* **40**, 10–19 (1993).
- Meece, D. *et al.* in *Preliminary Depth-age Scale of the GISP2 Ice Core* (Spec. Re. 94-1, Cold Regions Research and Engineering Lab., Hanover, New Hampshire, 1994).
- Mayewski, P. A. *et al.* Changes in atmospheric circulation and ocean ice cover over the North Atlantic during the last 41,000 years. *Science* **263**, 1747–1751 (1994).
- Brook, E. J., Sowers, T. & Orchardo, J. Rapid variations in atmospheric methane concentration during the past 110,000 years. *Science* **273**, 1087–1091 (1994).
- Bard, E., Rostek, F. & Sonzogni, C. Interhemispheric synchrony of the last deglaciation inferred from alkenone palaeothermometry. *Nature* **385**, 707–710 (1997).
- Stuiver, M. & Braziunas, T. F. Modeling atmospheric <sup>14</sup>C-influences and <sup>14</sup>C ages of marine samples to 10,000 BC. *Radiocarbon* **35**, 137–189 (1993).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank R. Behl, U. Berner, M. den Dulk, P. Grootes, L. Labeyrie, M. Maslin and R. Zahn for suggestions about earlier versions of the manuscript; M. Weber for processing the core logger sound velocity data of cores SO90-88/93KL; and T. Weiser, M. Geyh and P. Grootes for providing electron microprobe and radiocarbon data.

Correspondence and requests for materials should be addressed to H.S. (e-mail: hartmut.schulz@io-warnemuende.de).

# Re–Os isotope evidence for the composition, formation and age of the lower continental crust

A. E. Saal\*, R. L. Rudnick†, G. E. Ravizza\* & S. R. Hart\*

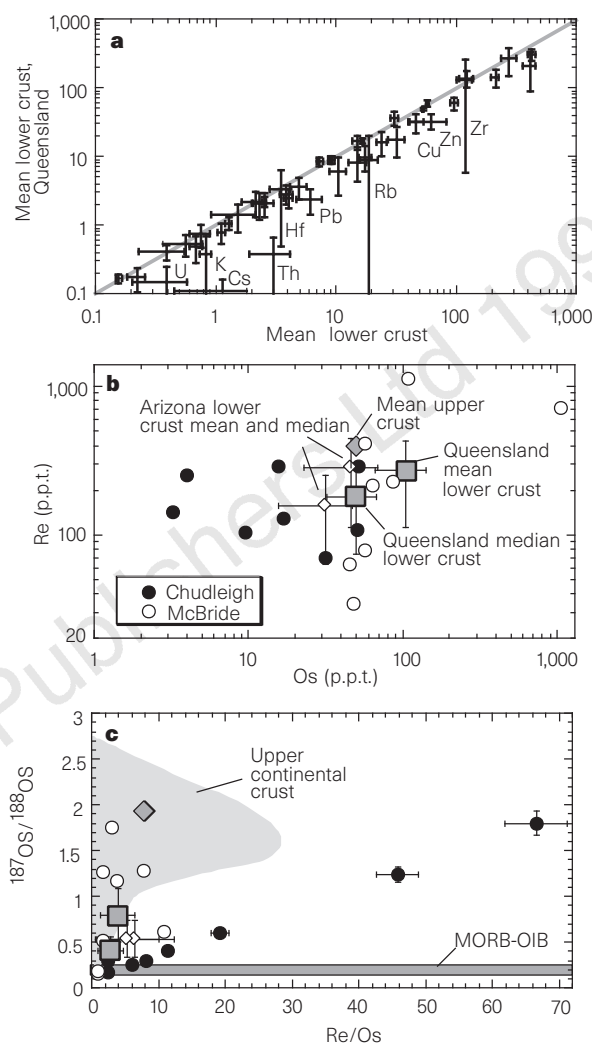
\* Department of Geology and Geophysics, Woods Hole Oceanographic Institute, Woods Hole, Massachusetts 02543, USA

† Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA

Knowledge of the composition of the lower continental crust is important for understanding the formation and evolution of the crust as a whole, and the petrogenesis of continental basalts. Here we present rhenium–osmium isotope data for two well characterized suites of lower-crustal xenoliths from North Queensland, Australia<sup>1–7</sup>, which have average major- and trace-element compositions similar to estimates of the bulk lower continental crust<sup>8–9</sup>. Our data indicate that the lower crust has 1 to 2 times as much osmium, about half as much rhenium, and is less radiogenic than the upper continental crust<sup>10</sup>. We interpret the rhenium–osmium isotope systematics to indicate that assimilation and fractional crystallization are important processes in the formation of the lower crust, and lead to dramatic changes in the osmium isotopic composition of basalts that pond and fractionate there. A consequence of this is that the rhenium–osmium isotopic system should not be relied on to yield accurate mantle extraction ages for continental rocks.

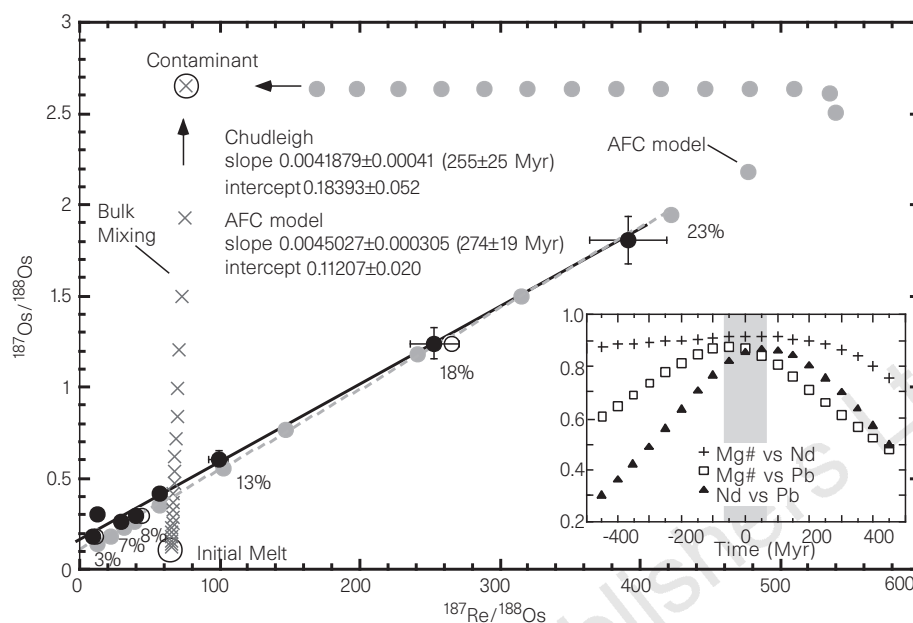
Re–Os differs from other isotopic systems (Pb, Sr and Nd) because Re and Os are siderophile and chalcophile (rather than lithophile) elements, and because Os is a compatible element (Os is retained in the crystals) during partial melting or fractional crystallization<sup>11</sup>. The fact that Os is a compatible element leads to high Re/Os ratios and radiogenic Os isotopic compositions in the continental crust<sup>12</sup> and makes <sup>187</sup>Os/<sup>188</sup>Os ratios in fractionating mantle-derived magmas susceptible to change by crustal contamination<sup>13–18</sup>.

The Os isotopic composition and Re and Os concentrations of 16 samples from two well characterized lower-crustal xenolith suites from North Queensland, Australia<sup>1–7</sup>, are reported in Table 1. These xenoliths have average major- and trace-element compositions similar to estimates of the bulk lower continental crust<sup>8,9</sup> (Fig. 1a), and can therefore be used to place broad constraints on its Re–Os concentration and Os isotopic composition. The xenoliths have average Os and Re concentrations of respectively 102 and 276 parts per 10<sup>12</sup> (p.p.t.; Re/Os = 2.71), with median values of 49 and 184 p.p.t. (Re/Os = 3.75), respectively (Table 1, Fig. 1b). The differences between the average and the median are produced by one outlier (sample 85–100) with high Os (~1,000 p.p.t.) and Re (~800 p.p.t.) concentrations. The median values are probably more representative of the bulk lower crust, but more data are clearly desirable. The mean, concentration-weighted, <sup>187</sup>Os/<sup>188</sup>Os ratio for the lower crust in north Queensland is 0.4 or 0.8, including and excluding the outlier, respectively. Recent analysis of eclogitic and granulite facies xenoliths from central Arizona give similar results (Table 1, Fig. 1b and c)<sup>19</sup>, although these have somewhat lower Os contents which translate into a slightly higher mean Re/Os ratio. We infer from these data that, relative to the upper crust<sup>10</sup>, the lower crust has one to two times as much Os, about half of the Re, and is probably less radiogenic (having one-half to one-third the <sup>187</sup>Os/<sup>188</sup>Os of the upper crust). However, the Os isotopic composition of the lower



**Figure 1** Representativeness and Re–Os composition of Queensland lower-crustal xenoliths. **a**, Comparison between the mean major- and trace-element composition of global lower-crustal xenoliths with those from the Queensland xenoliths. The global dataset (available on <http://www-ep.es.lnl.gov/germ/>) includes 450 analyses of major elements and 150 analyses of trace elements, excluding the Queensland data. Both means agree within the standard errors of the compilations, with the exception of Cs, Th, U, Pb, Cu and Zn, which we attribute to the mobility of these elements during secondary processes and the difficulty in their accurate measurement. The fact that many lower-crustal xenoliths are demonstrably older than the magmatic episode that carried them to the surface<sup>8</sup> suggests that the global average is a reasonable estimate of lower-crust composition. **b**, Os versus Re concentrations, and **c**, Re/Os versus <sup>187</sup>Os/<sup>188</sup>Os ratios for lower-crustal xenoliths from Queensland, Australia. The two mean lower-crustal isotopic compositions shown in **c** are weighted according to Os concentrations (filled squares). The lower one includes an anomalously high Os point, the upper one excludes this point. Fields for upper continental crust (without organic and carbonate-shelf sediments) (filled diamond) and oceanic basalts (mid-ocean ridge basalts and oceanic island basalts) were compiled by B. Peucker-Ehrenbrink. The mean and median Re and Os concentration, Re/Os ratios and isotopic composition weighted by Os concentration for Arizona xenoliths<sup>19</sup> (open diamonds) are plotted for comparison. The McBride xenoliths (open circles) show chemical characteristics indicative of a lithologically and genetically diverse suite, and plot in the field defined by upper-crustal material. The Chudleigh suite (filled circles) shows characteristics of genetically related cumulates from basaltic magma undergoing crustal contamination (see text).





**Figure 2** Re–Os isochron diagram for the Chudleigh suite showing bulk mixing and AFC trends. The data define a linear trend (filled circles), defining an apparent age of ~260 Myr. Inset shows correlation coefficients (for linear correlations) between Mg#, Nd ( $^{143}\text{Nd}/^{144}\text{Nd}$ ) and Pb ( $^{206}\text{Pb}/^{204}\text{Pb}$ ). In contrast to the Re–Os age, the best correlations exist at < 100 Myr, suggesting that the crystallization age of these xenoliths was < 100 Myr. The AFC model (using the parameters listed in Fig. 3 legend) generates a linear trend in the Re–Os diagram (shown by grey circles), which has no age significance. Open circles are duplicate analyses, not included in the isochron regression. X's show the bulk mixing trend between the

initial melt and the assimilant used in the AFC model. Each grey circle and X corresponds to 5% AFC or bulk mixing, respectively. The York regression for the AFC model was calculated using the same number of points and same distribution in the isochron as the Chudleigh suite to make the results comparable (the points used are labelled; for example 3%). We note that the Chudleigh xenoliths are cumulates, producing low Re/Os during the initial stages of AFC. Figs 2 and 3 indicate that careful evaluation of AFC is necessary before age significance is attributed to Re–Os isochrons for lower-crustal xenoliths or for basalts (or their cumulates) that may have differentiated in the lower continental crust.

**Table 1** Re and Os concentrations and Os isotopic compositions of granulite-facies lower-crustal xenolith suites from Northern Queensland, Australia

| Sample                               | Lithology         | Re (p.p.t.) | Os (p.p.t.) | $^{187}\text{Re}/^{188}\text{Os}$ | $^{187}\text{Os}/^{188}\text{Os}$ | $T_{\text{Os,CH}}$ (Myr) |
|--------------------------------------|-------------------|-------------|-------------|-----------------------------------|-----------------------------------|--------------------------|
| Chudleigh suite                      |                   |             |             |                                   |                                   |                          |
| 83-107                               | Pl-rich cumulate  | 153.2       | 3.20        | 265                               | 1.2492                            | 258                      |
|                                      |                   | 146         |             | 252                               |                                   | 271                      |
| 83-114*                              | Pl-rich cumulate  | 109.6       | 49.7        | 10.66                             | 0.18750                           | 349                      |
|                                      |                   | 89.8        | 53.8        | 8.08                              | 0.18895                           | 478                      |
| 83-127                               | Pl-rich cumulate  | 132.2       | 16.5        | 39.5                              | 0.30557                           | 275                      |
|                                      |                   | 146.6       |             | 43.9                              |                                   | 248                      |
| 83-131                               | Pl-rich cumulate  | 104.7       | 9.39        | 56                                | 0.4251                            | 324                      |
| 83-140                               | Pl-rich cumulate  | 259         | 3.85        | 391                               | 1.8138                            | 263                      |
| 83-110                               | Px-rich cumulate  | 70.8        | 31.3        | 11.27                             | 0.31038                           | 1,013                    |
| 83-115                               | Px-rich cumulate  | 292         | 15.3        | 97.9                              | 0.6132                            | 302                      |
| BC                                   | Transitional cum. | 297         | 51          | 28.7                              | 0.27030                           | 305                      |
| McBride suite                        |                   |             |             |                                   |                                   |                          |
| 83-160                               | Felsic granulite  | 219         | 61.9        | 19.33                             | 1.1799                            | 3,299                    |
| 83-162                               | Felsic granulite  | 80.1        | 56.2        | 7.23                              | 0.5013                            | 3,250                    |
| 85-100*                              | Mafic granulite   | 734         | 1,043       | 3.41                              | 0.1794                            | 897                      |
|                                      |                   | 808         | 931         | 4.21                              | 0.16524                           | 586                      |
| 85-108*                              | Mafic granulite   | 1,141       | 106.5       | 55.0                              | 0.6230                            | 550                      |
|                                      |                   | 1,210       | 104.5       | 59.3                              | 0.6079                            | 494                      |
| 85-106                               | Mafic cumulate    | 50.8        | 47.4        | 5.26                              | 0.19930                           | 884                      |
|                                      |                   | 34.5        |             | 3.58                              |                                   | 1,349                    |
| 83-159                               | Mafic residue     | 422         | 55.9        | 41.9                              | 1.2959                            | 1,692                    |
| 85-114*                              | Mafic residue     | 234         | 84          | 16.28                             | 1.7597                            | 5,970                    |
|                                      |                   | 230         | 86          | 15.46                             | 1.6590                            | 5,910                    |
| 83-157                               | Metasediment      | 64.1        | 44.5        | 8.00                              | 1.2832                            | 8,649                    |
| Crustal averages                     |                   |             |             |                                   |                                   |                          |
| North Queensland mean                |                   | 276         | 102         | 13.54                             | 0.418                             | 1,330                    |
| <b>North Queensland median</b>       |                   | <b>184</b>  | <b>49</b>   | <b>19.70</b>                      | <b>0.804</b>                      | <b>2,100</b>             |
| Central Arizona mean <sup>13</sup>   |                   | 286         | 45          | 32.33                             | 0.548                             | 2,278                    |
| Central Arizona median <sup>14</sup> |                   | 161         | 31          | 26.42                             |                                   | 796                      |
| Average upper crust <sup>10</sup>    |                   | 400         | 50          | 47.62                             | 1.9256                            | 976                      |

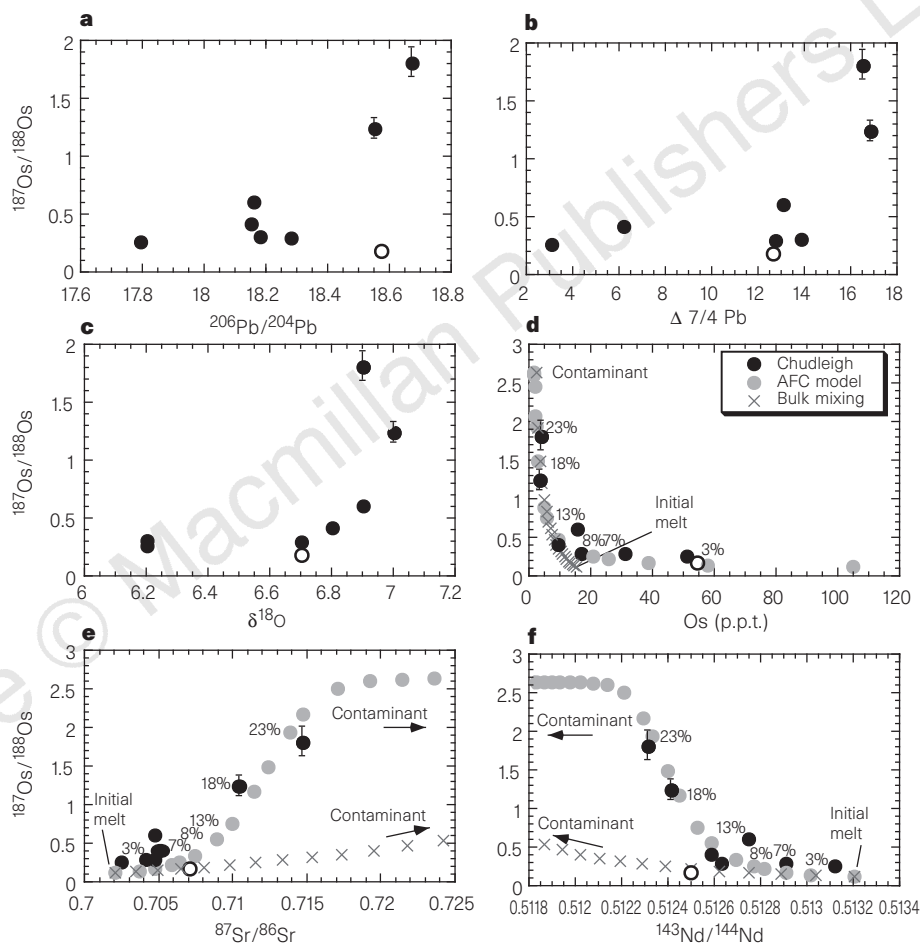
$T_{\text{Os,CH}}$  are Os model ages calculated with respect to 'bulk earth'  $^{187}\text{Re}/^{188}\text{Os} = 0.4243$  and  $^{187}\text{Os}/^{188}\text{Os} = 0.1287$  (ref. 21) and a decay constant of  $1.64 \times 10^{-11}$ ; Pl, plagioclase; Px, pyroxene. Analytical techniques are described in ref. 22. Run precision in all measurements is  $2\sigma < 0.35\%$  for  $^{187}\text{Os}/^{188}\text{Os}$ ,  $2\sigma < 1\%$  for Os and  $2\sigma < 2\%$  for Re contents (samples 83-107 and 83-140 have  $2\sigma < 10\%$  and  $< 7\%$  for Os concentration and isotopic composition, respectively, due to blank correction). The analytical blanks ranged from 4 to 4.7 for  $^{187}\text{Os}/^{188}\text{Os}$ , 0.4 to 0.5 pg g<sup>-1</sup> for Os and 5 to 24 pg g<sup>-1</sup> for Re. All samples were corrected for blank contribution. Re and Os were determined on separate powder splits subjected to acid digestion (Re) or fusion (Os). Repeat Re and Os concentration and  $^{187}\text{Os}/^{188}\text{Os}$  were made on separate powder splits. Os and Re replicates range from 2% to 11% and 1% to 18%, respectively (excluding 85-106 with 32% difference in the Re replicate).  $^{187}\text{Os}/^{188}\text{Os}$  ratios for replicates for the Chudleigh xenoliths are < 0.6%, but for the McBride suite, the replicates range from 2% to 5%. The style of the poor reproducibility in replicates in the McBride suite (that is, large variation in  $^{187}\text{Os}/^{188}\text{Os}$  ratios and Re contents, at similar Os concentrations) is suggestive of a Re 'nugget' effect—that is, domains or minerals having higher Re/Os are heterogeneously distributed within the rock, and have been out of diffusive equilibrium with the rest of the rock for reasonably long time spans (in this case ~300 Myr, the age of the suite<sup>3</sup>).

\* Indicates sample and replicate that were rerun on the same day.

continental crust will vary with the age of the crust. Given a mean age of 2.3 Gyr for the continental crust<sup>20</sup>, the Re–Os composition of the mantle<sup>21</sup> and our average Re/Os estimates including and excluding the data for the Arizona xenoliths<sup>19</sup> ( $^{187}\text{Re}/^{188}\text{Os} \approx 20$  and 16, respectively), we calculate the mean  $^{187}\text{Os}/^{188}\text{Os}$  of the bulk lower crust to be between 0.7 and 0.9 (Fig. 1c), a range that encompasses the median Os isotopic composition derived from the Queensland xenoliths (Table 1). Importantly, our data establish the presence of material with relatively high Os contents and radiogenic Os isotopic compositions in the lower crust. The variation in Os isotopic composition and Re–Os concentrations in all three xenolith suites illustrates the heterogeneity of the present

lower crust and shows that isotopic homogenization through fluid flow or melting has not been important.

The two suites of xenoliths from Queensland show differences in Re and Os concentration and Re/Os ratios, but have a similar, and large, range in  $^{187}\text{Os}/^{188}\text{Os}$  ratios (from 0.1721 and 1.759 versus 0.1889 to 1.814) (Fig. 1a and b). McBride xenoliths have a larger range of Re (two orders of magnitude), and a smaller range of Os contents and Re/Os ratios than Chudleigh xenoliths (a factor of less than two, excluding the outlier). These characteristics are consistent with the extreme lithologic and genetic diversity of McBride xenoliths, which range from paragneisses to mafic and felsic orthogneisses formed during large-scale mixing between Protero-



**Figure 3** Isotopic correlations and AFC model for Chudleigh xenoliths.  $^{187}\text{Os}/^{188}\text{Os}$  versus  $^{206}\text{Pb}/^{204}\text{Pb}$  (a),  $\Delta 7/4 \text{ Pb}$  (see ref. 23) (b),  $\delta^{18}\text{O}$  (c), Os concentration (d),  $^{143}\text{Nd}/^{144}\text{Nd}$  (e) and  $^{87}\text{Sr}/^{86}\text{Sr}$  (f) for the Chudleigh suite. Relatively good correlations exist between  $^{187}\text{Os}/^{188}\text{Os}$ , Os concentration and Pb, Sr, Nd, Pb, O isotopic systems, with the exception of one sample. The aberrant sample (83-114, open circle) is probably showing the 'nugget' effect (see Table 1); this explanation is consistent with this sample having the lowest Os isotopic ratios and the highest Os concentration, and plotting outside the correlation defined by the other samples in Os–Ni and Os–Mg# plots (not shown). The AFC curves are marked in 5% increments (except the labelled points), where the ratio of mass assimilated to mass fractionated is  $\approx 0.67$ . The contaminant has intermediate Sr and Nd isotopic compositions ( $^{87}\text{Sr}/^{86}\text{Sr} \approx 0.75$  and  $^{143}\text{Nd}/^{144}\text{Nd} \approx 0.5114$ ) and elemental concentrations (Sr  $\approx 200$  p.p.m., Nd  $\approx 27$  p.p.m.), based on published data for Tasman fold belt crust<sup>1</sup>. It is assumed to have very low Os ( $\sim 2$  p.p.t.) to test the sensitivity of Os to crustal contamination, and Re concentrations ( $\sim 24$  p.p.t.), and  $^{87}\text{Os}/^{188}\text{Os} \approx 2.6477$  ( $T_{\text{OsCH}} \approx 2$  Gyr; see Table 1 for definition), similar to Proterozoic upper continental

crust. All isotope systems (for example, Sr, Nd, Pb, O and Os) require a contaminant that is more evolved than any sample seen in either the Chudleigh or McBride xenolith suites and suggests the presence in the lower crust of (unsampled) old material with high  $\delta^{18}\text{O}$  and time-integrated high Rb/Sr, U/Pb and Re/Os, and low Sm/Nd. For the parental basaltic melt we used the most primitive isotopic composition seen in the xenoliths ( $^{87}\text{Sr}/^{86}\text{Sr} \approx 0.702$ ,  $^{143}\text{Nd}/^{144}\text{Nd} \approx 0.5131$ ) and inferred  $^{187}\text{Os}/^{188}\text{Os} \approx 0.1276$ . We assumed that this melt has Sr  $\approx 350$  p.p.m., Nd  $\approx 14$  p.p.m., and has Re  $\approx 250$  p.p.t., and 7.5 times the Os concentration of the contaminant ( $\sim 15$  p.p.t.). Bulk partition coefficients for Sr and Nd were estimated assuming cumulates of plagioclase + clinopyroxene + olivine ( $D^{\text{Sr}} \approx 0.5$ ,  $D^{\text{Nd}} \approx 0.1$ ). The bulk partition coefficient for Os ( $D^{\text{Os}} \approx 7$ ), was estimated from the correlation of Os–Ni (bulk  $D^{\text{Ni}} \approx 7.4$ )<sup>11</sup>; Re was assumed to be moderately incompatible ( $D^{\text{Re}} \approx 0.6$ ). The AFC model presented here is not unique, but serves to illustrate that the trends observed in the data are explained by an AFC process; the same is not true for bulk mixing, which is shown for comparison.  $D$  is the bulk partition coefficient of an element between the solid crystals and the residual melt.

zoic crust and basalts during late Palaeozoic (~300 Myr age) convergent margin magmatism<sup>2-7</sup>. Chudleigh xenoliths are mafic in composition ( $\text{SiO}_2 < 51.2 \text{ wt}\%$ ), showing good correlations between Sr, Nd, Pb and O isotopic composition, trace-element contents and Mg# (molar  $\text{Mg}/(\text{Mg} + \text{Fe})$  ratios)<sup>1,4,6,7</sup>. These correlations reflect a genetically related suite of cumulates from mafic magma(s) that underwent simultaneous assimilation and fractionation in the deep crust<sup>1</sup>. It is very unusual to find such simple chemical and isotopic systematics in lower-crustal xenolith suites. The Chudleigh suite thus provides a unique natural laboratory with which to investigate the chemical changes occurring to mafic magmas that pond and fractionate within the deep continental crust.

Although the Chudleigh xenoliths cannot be dated directly, the chemical systematics described above provide some constraints on their magmatic crystallization age. Because the xenoliths are cumulates, and not crystallized magmas, they have a large range in Sm/Nd ratio, depending upon whether plagioclase or clinopyroxene is the main cumulate phase. Thus the excellent correlations between Nd isotopic composition (which ranges from  $\epsilon_{\text{Nd}}$  of +9.6 to -6.1), Mg# (63–79) and Pb ( $^{206}\text{Pb}/^{204}\text{Pb} = 17.8\text{--}18.7$ ), Sr (0.70239–0.71467) and O ( $\delta^{18}\text{O} = +6.2$  to  $+7.0$ ) isotopes seen today degrade as the isotopic compositions are calculated at earlier times<sup>1,4</sup> (Fig. 2, inset). These features led to the conclusion that the cumulates crystallized relatively recently (<100 Myr ago), and that they are probably related to Cenozoic igneous activity that occurs throughout eastern Australia<sup>1</sup>.

The Re–Os results for the Chudleigh xenoliths show a strikingly good linear correlation on a Re–Os isochron diagram, defining an apparent age of ~260 Myr (Fig. 2). There are two possible interpretations of this correlation: (1) it accurately reflects the igneous crystallization age of the xenoliths, and (2) it reflects recent mixing between a mantle-like endmember and a radiogenic crustal endmember. We prefer the second interpretation for the reasons outlined below.

Parental magmas to the xenoliths underwent both crystallization (as evidenced by the cumulate xenoliths) and crustal assimilation (as seen in the Nd, Sr, Pb, O isotopic systems, trace-element and Mg# systematics)<sup>1,4,7</sup>. If interpretation (1) is correct, at ~260 Myr ago either Os isotopes were completely immune to crustal assimilation, and the samples should all have the same initial  $^{187}\text{Os}/^{188}\text{Os}$ , or Os was affected and should therefore correlate with other indicators of contamination (for example, O isotopes). Neither is true. The ( $^{187}\text{Os}/^{188}\text{Os}$ )<sub>260 Myr</sub> ratios of these xenoliths have a large range, between 0.12 and 0.26 (compare with 0.1276 to 0.1565 for modern mantle-derived basalts), and do not correlate with any of the other isotopic systems at ~260 Myr. In contrast, relatively good positive correlations are seen for present-day  $^{187}\text{Os}/^{188}\text{Os}$  and O, Pb and Sr isotopic compositions and negative correlations between  $^{187}\text{Os}/^{188}\text{Os}$  and Nd isotopic ratios (Fig. 3). The inset to Fig. 2 shows that the correlations between different isotope systems, as measured by their correlation coefficient, are best within 50–100 Ma of the present. The general lack of correlations between Mg# and Nd, Pb isotopes at 260 Myr (refs 1, 4; Fig. 2, inset), coupled with the relatively good present-day correlations<sup>1,4</sup> (Fig. 3), strongly suggests that these cumulates did not crystallize at 260 Myr. We conclude that the positive trend in the Re–Os isochron diagram was produced by assimilation and fractional crystallization (AFC) processes at a later date.

We model the Chudleigh xenoliths as cumulates resulting from simultaneous AFC of a basaltic melt in the deep continental crust (Fig. 3d–f; modelling parameters are given in Fig. 3 legend). The sigmoidal shape of the curves shows that during the first stages of AFC the Os isotopic composition is buffered by the Os content in the basaltic magma, whereas during the later stages it is buffered by the Os content and isotopic composition of the assimilate (old continental crust). The fractionation of olivine and/or sulphide

decreases the concentration of Os in the magma<sup>11</sup>, making it susceptible to contamination with crustal material having low Os concentration (2 p.p.t., in our model). We note that in the initial and final stages there is decoupling between  $^{187}\text{Os}/^{188}\text{Os}$  and the other isotopic systems. The compatible behaviour of Os relative to Nd, Sr and Pb, and the differences in Os concentration and isotopic composition between mantle melts and evolved crustal materials, accounts for the buffering of Os in the melt; such effects are generally not observed in the other isotopic systems. Moreover, the compatible character of Os explains the large difference between the trends produced by AFC and bulk mixing (Figs 2 and 3) and makes AFC a far more effective process for changing the isotopic composition of the basalt than bulk mixing, especially when the contaminant has a low Os concentration. Not only does the AFC model explain the correlation between  $^{187}\text{Os}/^{188}\text{Os}$  and the other isotopic ratios observed in the Chudleigh xenoliths (Fig. 3d–f), it also reproduces the near-linear trend in the Re–Os isochron diagram (Fig. 2). Our data show that apparent whole-rock Re–Os isochrons with no age significance may be produced during AFC processes by depletion of Os through olivine-sulphide fractionation, coupled with assimilation of radiogenic crust. Moreover, small amounts (<20%) of AFC during basalt/lower-crust interaction can produce isotopic variations in the melt that could be erroneously interpreted as evidence for mantle heterogeneity. □

Received 10 October 1997; accepted 17 February 1998.

- Rudnick, R. L., McDonough, W. F., McCulloch, M. T. & Taylor, S. R. Lower crustal xenoliths from Queensland, Australia: evidence for deep crustal assimilation and fractionation of continental basalts. *Geochim. Cosmochim. Acta* **50**, 1099–1115 (1986).
- Rudnick, R. L. & Taylor, S. R. The composition and petrogenesis of the lower crust: a xenolith study. *J. Geophys. Res.* **92**(B13), 13981–14005 (1987).
- Rudnick, R. L. & Williams, I. S. Dating the lower crust by ion microprobe. *Earth Planet. Sci. Lett.* **85**, 145–161 (1987).
- Rudnick, R. L. & Goldstein, S. L. The Pb isotopic compositions of lower crustal xenoliths and the evolution of the lower crustal Pb. *Earth Planet. Sci. Lett.* **98**, 192–207 (1990).
- Rudnick, R. L. Nd and Sr isotopic compositions of lower crustal xenoliths from North Queensland, Australia: implications for Nd model ages and crustal growth processes. *Chem. Geol.* **83**, 195–208 (1990).
- Rudnick, R. L. & Taylor, S. R. in *The Australian Lithosphere* (ed. Drummond, B.) 189–208 (Spec. Publ. 17, Geol. Soc. Aust., Canberra, 1991).
- Kempton, P. D. & Harmon, R. S. Oxygen isotope evidence for large-scale hybridization of the lower crust during magmatic underplating. *Geochim. Cosmochim. Acta* **56**, 971–986 (1992).
- Rudnick, R. L. in *The Continental Lower Crust* (eds Fountain, D. M., Arculus, R. & Kay, R.) 269–316 (Elsevier, Amsterdam, 1992).
- Rudnick, R. L. & Fountain, D. M. Nature and composition of the continental crust: a lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
- Esser, B. K. & Turekian, K. K. The osmium isotopic composition of the continental crust. *Geochim. Cosmochim. Acta* **57**, 3093–3104 (1993).
- Hart, S. R. & Ravizza, G. E. in *Earth Processes: Reading the Isotopic Code* (eds Basu, A. & Hart, S. R.) 123–134 (AGU Monogr. 95, Am. Geophys. Union, Washington DC, 1996).
- Allègre, C. J. & Luck, J. M. Osmium isotopes as petrogenetic and geological tracers. *Earth Planet. Sci. Lett.* **48**, 148–154 (1980).
- Reisberg, L. C. *et al.* Os isotope systematics in ocean island basalts. *Earth Planet. Sci. Lett.* **120**, 149–167 (1993).
- Marcantonio, F., Zindler, A., Elliot, T. R. & Staudigel, H. Os isotope systematics of La Palma, Canary Islands: evidence for recycled crust in the mantle source of HIMU ocean island. *Earth Planet. Sci. Lett.* **133**, 397–410 (1995).
- Roy-Barman, M. & Allègre, C. J.  $^{187}\text{Os}/^{188}\text{Os}$  in oceanic island basalts: tracing crust recycling in the mantle. *Earth Planet. Sci. Lett.* **129**, 145–161 (1995).
- Widom, E. & Shirey, S. B. Os isotope systematics in the Azores: implications for mantle plume sources. *Earth Planet. Sci. Lett.* **142**, 451–465 (1991).
- Martin, C. E., Carlson, R. W., Shirey, S. B., Frey, F. A. & Chen, C. Y. Os isotopic variation in basalts from Haleakala volcano, Maui, Hawaii: a record of magmatic processes in oceanic mantle and crust. *Earth Planet. Sci. Lett.* **128**, 287–301 (1991).
- Walker, R. J., Morgan, J. W., Hanski, E. J. & Smolkin, V. F. Re–Os systematics of early Proterozoic ferroporphyrates, Pechenga Complex, NW Russia: evidence for ancient  $^{187}\text{Os}$ -enriched plumes. *Geochim. Cosmochim. Acta* **61**, 3145–3160 (1997).
- Esperança, S., Carlson, R. W. & Shirey, S. B. Dating crust–mantle separation: Re–Os isotopic study of mafic xenoliths from central Arizona. *Geology* **25**, 651–654 (1997).
- Allègre, C. J., Hart, S. R. & Minster, J. F. Chemical structure and evolution of the mantle and continents determined by inversion of Nd and Sr isotope data II. Numerical experiments and discussion. *Earth Planet. Sci. Lett.* **66**, 191–213 (1983).
- Meisel, T., Walker, R. J. & Morgan, J. W. The osmium isotopic composition of the Earth's primitive upper mantle. *Nature* **383**, 517–520 (1996).
- Peucker-Ehrenbrink, B., Ravizza, G. E. & Hofmann, A. W. The marine  $^{187}\text{Os}/^{186}\text{Os}$  record of the past 80 million years. *Earth Planet. Sci. Lett.* **130**, 155–167 (1995).
- Hart, S. R. A large-scale isotopic anomaly in the Southern Hemisphere mantle. *Nature* **309**, 753–757 (1984).

**Acknowledgements.** We thank J. T. Chesley, W. F. McDonough and D. Pyle for comments on the manuscript, and R. Carlson for comments that improved our presentation. We also thank B. Peucker-Ehrenbrink for making accessible his Os data set.

Correspondence and requests for materials should be addressed to A.S. (e-mail: saal@whoi.edu).

# New specimens and confirmation of an early age for *Australopithecus anamensis*

Meave G. Leakey\*, Craig S. Feibel†, Ian McDougall‡, Carol Ward§ & Alan Walker||

\* National Museums of Kenya, PO Box 40658, Nairobi, Kenya

† Department of Anthropology, Rutgers University, New Brunswick, New Jersey 08901, USA

‡ Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia

§ Departments of Anthropology and Pathology & Anatomical Sciences, University of Missouri, Columbia, Missouri 65211, USA

|| Departments of Anthropology and Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

The discovery of *Australopithecus anamensis* fossils<sup>1</sup> from strata lying between tephra dated at 4.17 and 4.12 million years ago, and from slightly higher strata not well constrained in age by overlying dated units, provoked the claim that more than one species might be represented: it was suggested that the stratigraphically higher fossils, which include the important tibia, humerus and a large, presumed male, mandible (KNM-KP 29287), might belong to a later, more derived hominid<sup>2</sup>. We have recovered new fossils from Kanapoi and Allia Bay, Kenya, during field work in 1995–1997 that confirm the primitive status of *Australopithecus anamensis*, the earliest species of *Australopithecus*. Isotope dating confirms *A. anamensis*' intermediate age as being between those of *Ardipithecus ramidus*<sup>3,4</sup> and *Australopithecus afarensis*<sup>5,6</sup>. New specimens of maxilla, mandible and capitate show that this species is demonstrably more primitive than *A. afarensis*. A lower first deciduous molar (dm1) is intermediate in morphology between that reported for *Ardipithecus ramidus*<sup>4</sup> and *A. afarensis*<sup>7</sup>. Single-crystal <sup>40</sup>Ar–<sup>39</sup>Ar age determinations on the Kanapoi Tuff show that, except for a large mandible, all of the hominid fossils from Kanapoi are from sediments deposited between  $4.17 \pm 0.03$  and  $4.07 \pm 0.02$  million years ago.

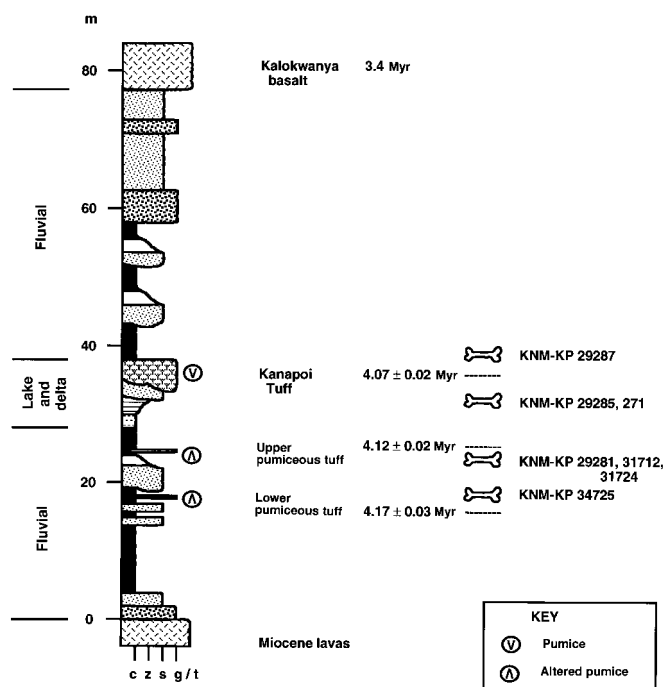
The stratigraphic sequence at Kanapoi has three major lithologic intervals (Fig. 1). (1) A basal fluvial complex of channel sandstones and floodplain palaeosols—most of the hominids come from this well dated interval. (2) Claystones with ostracods and molluscs, which we interpret as being deposited in the early Lonyumun Lake, well represented throughout the Turkana basin<sup>8</sup>. The upper strata of this interval include extensive deltaic sandstones which are rich in vertebrate remains, including those of *A. anamensis*. (3) A second fluvial interval, with extensive conglomeratic channels near its top and capped by the Kalokwanya basalt.

Within the upper portion of the lacustrine sequence (layer 2) is an extensively exposed tuff, typically fine-grained and very light grey (N8) to yellowish grey (5Y 7/2) in colour. This tuff reaches more than 11 m in thickness, and contains planar, ripple and climbing-ripple laminations, together with extensive contorted bedding in its upper few decimetres. Basal topography indicates that it locally fills distributary channels, and it has been suggested<sup>9</sup> that much of the deposit reflects an interdistributary bay fill. Good exposures of this tuff are found around Akai Ititi, in the central part of the Kanapoi region. The individual glass shards show stretched and bubble-wall morphologies and, when cleaned of adhering clays, have a characteristic olive (5Y 5/3) colour. We designate this tephra the Kanapoi Tuff, with sample K92-4847 (Table 1) as its type example. Analysis of the type and of reference samples from nearby exposures

(K92-4846) and from the original collection<sup>9</sup> (KP01-15-01) show the tephra has a high-iron rhyolitic composition (Table 1). The high iron content is outside the range of reported values for Omo Group tephra from the northern part of the Turkana Basin<sup>10</sup>, and supports evidence from sedimentary patterns and geochemical correlation that this tephra may have been derived from a volcanic source to the south, towards the Baringo Basin.

Concentrations of small pumices can be found locally in thin lenses within the upper one to two metres of the Kanapoi Tuff. Sample 95-162 comprises small round to slightly flattened pumices ranging in size from about 5 to 30 mm diameter collected from a lens within the Kanapoi Tuff about 500 m south-southwest of Akai Ititi. Alkali feldspar phenocrysts were separated by heavy liquids, with subsequent hand-picking from the composite sample of these small pumices. Similarly, alkali feldspar was separated from two larger individual pumices (95-168 and 95-170A) collected from localities about 500 m east-southeast and east of Akai Ititi, respectively. These pumices contain sporadic crystals of alkali feldspar up to about 1.5 mm in size, which are ideal for age determination by the <sup>40</sup>Ar–<sup>39</sup>Ar method. Before dating, the alkali feldspar crystals were ultrasonically cleaned in 7% hydrofluoric acid for five minutes to remove traces of adhering matrix and volcanic glass.

Analytical results for the alkali feldspars from the three samples are listed in Table 2. Feldspars from the composite sample 95-162 yielded a concordant set of ages on nine crystals with a simple mean of  $4.072 \pm 0.019$  Myr, where the uncertainty is the standard deviation of the population. Measured ages from seven alkali feldspar crystals from pumice 95-168 also gave a concordant data set of mean age  $4.063 \pm 0.031$  Myr. Likewise, results from six crystals from pumice 95-170A are concordant and yield an average age of  $4.076 \pm 0.018$  Myr. As the mean ages for the three samples are statistically indistinguishable, we can combine results for the 22 separate single crystal analyses to yield a pooled mean age of  $4.070 \pm 0.023$  Myr, or  $\pm 0.005$  Myr if we calculate the standard error. If we use a weighted mean age using all the single-crystal measurements, we obtain a mean age of  $4.071 \pm 0.004$  Myr from an essentially gaussian distribution. As the error in the age of the fluence monitor is not well known, but is at least 0.5%, it is



**Figure 1** Composite section of strata at Kanapoi showing relationship of major hominid fossils to dated levels. Lithological abbreviations on column are: c, clay; z, silt; s, sand; g/t, gravel/tuff.

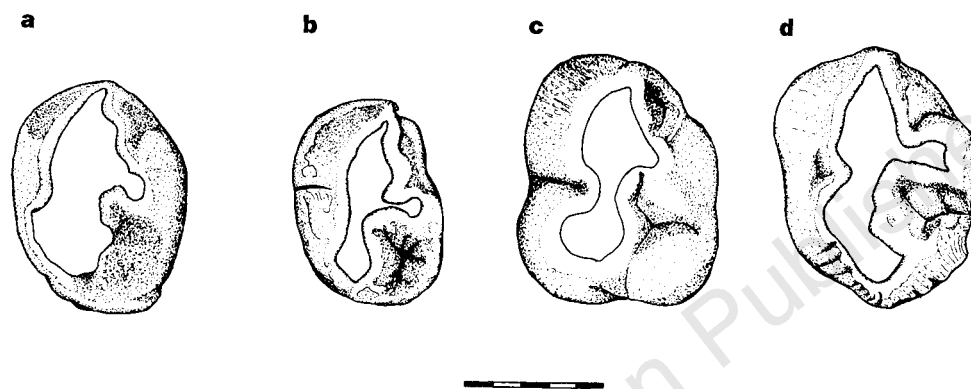


unrealistic and even misleading to quote errors much less than this. We therefore believe that the simple average age of  $4.070 \pm 0.023$  Myr is the best estimate we can make for the age of cooling of the alkali feldspars in the pumices of the Kanapoi Tuff. The timing of eruption that produced the material comprising the Kanapoi Tuff is also likely to closely approximate the timing of deposition of the tuff. Compared with the uncertainty in the measured age or the time elapsed subsequently, such materials blanketing a landscape are likely to be washed into depositional basins within geologically short intervals, of say a few thousand years.

With the dating of the Kanapoi Tuff, we have shown that all but

one of the Pliocene hominid fossils found in the Kanapoi sequence are well constrained in age to a small time interval between  $4.17 \pm 0.03$  Myr and  $4.07 \pm 0.02$  Myr. The tibia, coming from a distributary channel sandstone 4–8 m below the Kanapoi Tuff and the humerus which, from documentary evidence, comes from about the same level, are therefore only a little younger than the rest of the hominids. The mandible KNM-KP 29287 comes from a palaeosol immediately above the lacustrine deposits and is likely to be only marginally younger than the Kanapoi Tuff.

Several more specimens of *A. anamensis* have been recovered from Kanapoi and the slightly younger site of Allia Bay since our original report<sup>1</sup>. These are listed in Table 3. The most important for



**Figure 2** Drawings of lower first deciduous molars (dm<sub>1</sub>). **a**, *A. anamensis* KNM-KP 34725; **b**, *A. ramidus* ARA-VP-1/129; and **c**, **d**, *A. afarensis* LH2 and AL333 43b, artificially worn to the condition seen in **a** to show the intermediate morphology of this tooth. Some figures have been reversed for easy comparison; **b-d** were drawn from casts. Scale is in mm.

**Table 1** Electron microprobe analysis of glass from the Kanapoi Tuff

| Sample     | SiO <sub>2</sub> | Al <sub>2</sub> O <sub>3</sub> | Fe <sub>2</sub> O <sub>3</sub> | CaO  | K <sub>2</sub> O | Na <sub>2</sub> O | MgO  | MnO  | TiO <sub>2</sub> | Cl   | F    | Zr   | Total | N  |
|------------|------------------|--------------------------------|--------------------------------|------|------------------|-------------------|------|------|------------------|------|------|------|-------|----|
| KP01-15-01 | 70.36            | 7.86                           | 8.32                           | 0.37 | 3.93             | 1.73              | 0.01 | 0.25 | 0.25             | 0.42 | 0.03 | NA   | 93.53 | 6  |
| K92-4846   | 69.50            | 7.55                           | 8.31                           | 0.28 | 0.26             | 0.16              | 0.01 | 0.25 | 0.23             | 0.52 | 0.01 | 0.29 | 96.49 | 13 |
| K92-4847   | 69.76            | 7.66                           | 8.42                           | 0.29 | 0.37             | 0.24              | 0.01 | 0.25 | 0.24             | 0.48 | 0.01 | 0.28 | 96.58 | 18 |

Abundances are shown as weight per cent. *N*, number of shards analysed. NA, not analysed.

**Table 2** Results of <sup>40</sup>Ar–<sup>39</sup>Ar dating of single crystals from three localities in the Kanapoi Tuff, Kanapoi, Turkana, northern Kenya

| Sample                                                                                                                                                      | Mass (mg) | <sup>36</sup> Ar/ <sup>39</sup> Ar<br>× 10 <sup>-4</sup> | <sup>37</sup> Ar/ <sup>39</sup> Ar<br>× 10 <sup>-4</sup> | <sup>40</sup> Ar/ <sup>39</sup> Ar | K/Ca   | <sup>40</sup> Ar* / <sup>39</sup> Ar <sub>K</sub><br>± c.v. | <sup>40</sup> Ar* % | <sup>39</sup> Ar*<br>10 <sup>-14</sup> mol | Calculated age<br>Myr ± 1 s.d. |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|----------------------------------------------------------|----------------------------------------------------------|------------------------------------|--------|-------------------------------------------------------------|---------------------|--------------------------------------------|--------------------------------|
| 95-162 alkali feldspar from small pumices in tuff ~500 m SSW of Akai Ititi, 2° 18' 45" N, 36° 03' 42" E <i>J</i> = 0.001125 ± 0.35%; level 3, ANU23         |           |                                                          |                                                          |                                    |        |                                                             |                     |                                            |                                |
| Crystal 1                                                                                                                                                   | 2.8       | 5.331                                                    | 2.550                                                    | 2.175                              | 2064   | 1.996 ± 0.63%                                               | 91.8                | 4.91                                       | 4.046 ± 0.029                  |
| 2                                                                                                                                                           | 2.8       | 3.168                                                    | 0.030                                                    | 2.124                              | 176830 | 2.009 ± 0.65%                                               | 94.6                | 5.64                                       | 4.073 ± 0.030                  |
| 3                                                                                                                                                           | 1.7       | 2.926                                                    | 2.422                                                    | 2.126                              | 3173   | 2.018 ± 0.41%                                               | 94.9                | 3.08                                       | 4.091 ± 0.022                  |
| 4                                                                                                                                                           | 1.0       | 0.012                                                    | 1.223                                                    | 2.048                              | 4302   | 2.026 ± 0.80%                                               | 98.9                | 1.79                                       | 4.107 ± 0.036                  |
| 5                                                                                                                                                           | 0.8       | 2.849                                                    | 0.146                                                    | 2.103                              | 36110  | 1.998 ± 0.66%                                               | 95.0                | 1.39                                       | 4.050 ± 0.030                  |
| 6                                                                                                                                                           | 0.7       | 4.402                                                    | 0.732                                                    | 2.163                              | 7191   | 2.012 ± 0.45%                                               | 93.0                | 1.54                                       | 4.078 ± 0.023                  |
| 7                                                                                                                                                           | 3.9       | 5.153                                                    | 2.240                                                    | 2.184                              | 2350   | 2.010 ± 0.42%                                               | 92.0                | 7.34                                       | 4.075 ± 0.022                  |
| 8                                                                                                                                                           | 0.7       | 4.446                                                    | 2.801                                                    | 2.154                              | 1879   | 2.001 ± 0.38%                                               | 92.9                | 1.49                                       | 4.057 ± 0.021                  |
| 9                                                                                                                                                           | 1.5       | 3.096                                                    | 9.688                                                    | 2.123                              | 543    | 2.010 ± 0.35%                                               | 94.7                | 2.56                                       | 4.074 ± 0.020                  |
| $\bar{x}_9$                                                                                                                                                 |           |                                                          |                                                          |                                    |        |                                                             |                     |                                            | 4.072 ± 0.019                  |
| 95-168 alkali feldspar from single 11 g pumice in tuff; ~500 m ESE of Akai Ititi, 2° 18' 57" N, 36° 04' 01" E <i>J</i> = 0.001122 ± 0.35%; level 4, ANU23   |           |                                                          |                                                          |                                    |        |                                                             |                     |                                            |                                |
| Crystal 1                                                                                                                                                   | 1.4       | 11.343                                                   | 2.403                                                    | 2.379                              | 2190   | 2.023 ± 0.38%                                               | 85.0                | 2.96                                       | 4.090 ± 0.021                  |
| 2                                                                                                                                                           | 1.2       | 1.822                                                    | 16.524                                                   | 2.084                              | 318    | 2.010 ± 0.43%                                               | 96.4                | 1.76                                       | 4.063 ± 0.022                  |
| 3                                                                                                                                                           | 1.1       | 2.313                                                    | 7.205                                                    | 2.102                              | 731    | 2.012 ± 0.36%                                               | 95.7                | 1.88                                       | 4.069 ± 0.020                  |
| 4                                                                                                                                                           | 0.8       | 1.307                                                    | 2.854                                                    | 2.085                              | 1845   | 2.026 ± 0.43%                                               | 97.1                | 0.67                                       | 4.095 ± 0.023                  |
| 5                                                                                                                                                           | 0.7       | 22.095                                                   | 6.763                                                    | 2.656                              | 778    | 1.982 ± 1.24%                                               | 74.6                | 1.37                                       | 4.007 ± 0.051                  |
| 6                                                                                                                                                           | 1.2       | 1.886                                                    | 5.045                                                    | 2.093                              | 1043   | 2.107 ± 0.35%                                               | 96.3                | 2.20                                       | 4.077 ± 0.020                  |
| 7                                                                                                                                                           | 0.8       | 19.851                                                   | 20.654                                                   | 2.604                              | 255    | 1.997 ± 0.36%                                               | 76.7                | 1.34                                       | 4.037 ± 0.020                  |
| $\bar{x}_7$                                                                                                                                                 |           |                                                          |                                                          |                                    |        |                                                             |                     |                                            | 4.063 ± 0.031                  |
| 95-170A alkali feldspar from single 20 g pumice in tuff; ~500 m east of Akai Ititi, 2° 19' 05" N, 36° 03' 56" E <i>J</i> = 0.001116 ± 0.35%; level 6, ANU23 |           |                                                          |                                                          |                                    |        |                                                             |                     |                                            |                                |
| Crystal 1                                                                                                                                                   | 3.6       | 2.063                                                    | 2.629                                                    | 2.122                              | 2002   | 2.038 ± 0.42%                                               | 96.0                | 7.13                                       | 4.100 ± 0.022                  |
| 2                                                                                                                                                           | 2.6       | 2.211                                                    | 2.266                                                    | 2.111                              | 2323   | 2.022 ± 0.59%                                               | 95.8                | 4.80                                       | 4.069 ± 0.028                  |
| 3                                                                                                                                                           | 1.3       | 1.888                                                    | 1.720                                                    | 2.103                              | 3061   | 2.024 ± 0.42%                                               | 96.2                | 2.74                                       | 4.072 ± 0.022                  |
| 4                                                                                                                                                           | 1.1       | 4.338                                                    | 4.537                                                    | 2.187                              | 1160   | 2.036 ± 1.01%                                               | 93.1                | 2.53                                       | 4.096 ± 0.044                  |
| 6                                                                                                                                                           | 1.3       | 2.289                                                    | 3.288                                                    | 2.110                              | 1601   | 2.019 ± 0.38%                                               | 95.7                | 2.65                                       | 4.062 ± 0.021                  |
| 7                                                                                                                                                           | 1.1       | 1.536                                                    | 1.416                                                    | 2.086                              | 3717   | 2.018 ± 0.39%                                               | 96.7                | 2.28                                       | 4.059 ± 0.021                  |
| $\bar{x}_6$                                                                                                                                                 |           |                                                          |                                                          |                                    |        |                                                             |                     |                                            | 4.076 ± 0.018                  |

$\lambda = 5.543 \times 10^{-10} \text{ yr}^{-1}$ , where  $\lambda$  is the decay constant for <sup>40</sup>K. The fluence monitor used was sanidine 92-176 from the Fish Canyon Tuff, Colorado, which has a nominal K–Ar age of 27.9 Myr (ref. 14). The correction factors are as follows: (<sup>36</sup>Ar/<sup>39</sup>Ar)<sub>ca</sub> =  $3.50 \times 10^{-4}$ ; (<sup>37</sup>Ar/<sup>39</sup>Ar)<sub>ca</sub> =  $7.86 \times 10^{-4}$ ; (<sup>40</sup>Ar/<sup>39</sup>Ar)<sub>K</sub> = 0.020 to 0.030, depending upon the position in the irradiation vessel. Samples were irradiated with 0.2 mm Cd shielding. Data are corrected for decay of <sup>37</sup>Ar and <sup>39</sup>Ar. The Daly collector yielded <sup>40</sup>Ar/<sup>36</sup>Ar ratios on atmospheric argon between  $297.2 \times 2.0$  and  $300.5 \pm 1.7$  over the course of these measurements. <sup>40</sup>Ar\* indicates radiogenic argon. Overall mean age  $\bar{x}_{22} = 4.070 \pm 0.023$  Myr.

**Table 3 *Australopithecus anamensis* specimens collected since 1994 (excluding tooth fragments)**

| Specimen                | Year    | Elements                                                                                                                                                                                                                                                                                                                                            | Recovery                | Discoverer                    | Measurements                                                                                                                                                                                                                                                                                                                                           |
|-------------------------|---------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|-------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Allia Bay KNM-ER</b> |         |                                                                                                                                                                                                                                                                                                                                                     |                         |                               |                                                                                                                                                                                                                                                                                                                                                        |
| 30731                   | 1995    | R <sub>c</sub>                                                                                                                                                                                                                                                                                                                                      | <i>in situ</i>          | excavators                    | R <sub>c</sub> : 8.3MD, 10.6BL                                                                                                                                                                                                                                                                                                                         |
| 30744                   | 1996    | worn LP <sub>4</sub>                                                                                                                                                                                                                                                                                                                                | <i>in situ</i>          | excavators                    | LP <sub>4</sub> : >7.8MD, >9.0BL                                                                                                                                                                                                                                                                                                                       |
| 30745                   | 1996    | L maxilla (fragment C,P <sup>3</sup> -M <sup>1</sup> , partial M <sup>2</sup> , M <sup>3</sup> )                                                                                                                                                                                                                                                    | <i>in situ</i>          | excavators                    | LP <sup>3</sup> : 9.9MD, 13.0BL; LP <sup>4</sup> : 8.8MD, 13.9BL; LM <sup>1</sup> : 12.2MD, 14.1BLa, 13.9BLp; LM <sup>3</sup> : 12.7MD, >14.5BL                                                                                                                                                                                                        |
| 30750                   | 1996    | R <sub>c</sub>                                                                                                                                                                                                                                                                                                                                      | <i>in situ</i>          | excavators                    | R <sub>c</sub> : 6.6MD, 9.2BL                                                                                                                                                                                                                                                                                                                          |
| 35228                   | 1997    | RP <sub>4</sub>                                                                                                                                                                                                                                                                                                                                     | <i>in situ</i>          | excavators                    | RP <sub>4</sub> : 7.4MD, 9.6BL                                                                                                                                                                                                                                                                                                                         |
| 35231                   | 1997    | RM <sup>1</sup> or M <sup>2</sup>                                                                                                                                                                                                                                                                                                                   | <i>in situ</i>          | excavators                    | RM <sup>1</sup> : 11.7MD, 12.5BLa, 11.9BLp                                                                                                                                                                                                                                                                                                             |
| 35232                   | 1997    | LM <sup>1</sup>                                                                                                                                                                                                                                                                                                                                     | <i>in situ</i>          | excavators                    | LM <sup>1</sup> : 12.9MD, >12.2BLp                                                                                                                                                                                                                                                                                                                     |
| 35233                   | 1997    | LM <sup>2</sup> with roots                                                                                                                                                                                                                                                                                                                          | <i>in situ</i>          | excavators                    | LM <sup>2</sup> : 14.8MD, 13.3BLa, 13.0BLp                                                                                                                                                                                                                                                                                                             |
| 35235                   | 1997    | RM <sup>2</sup>                                                                                                                                                                                                                                                                                                                                     | <i>in situ</i>          | excavators                    | RM <sup>2</sup> : >12.9MD                                                                                                                                                                                                                                                                                                                              |
| 35236                   | 1997    | LM <sup>3</sup>                                                                                                                                                                                                                                                                                                                                     | <i>in situ</i>          | excavators                    | LM <sup>3</sup> : 11.5MD, 13.0BL                                                                                                                                                                                                                                                                                                                       |
| 35238                   | 1997    | RM <sup>1</sup>                                                                                                                                                                                                                                                                                                                                     | <i>in situ</i>          | excavators                    | RM <sup>1</sup> : >10MD, 12.8BLa, 12.3BLp                                                                                                                                                                                                                                                                                                              |
| <b>Kanapoi KNM-KP</b>   |         |                                                                                                                                                                                                                                                                                                                                                     |                         |                               |                                                                                                                                                                                                                                                                                                                                                        |
| 29283                   | 1997    | worn LI <sup>2</sup> (associated with 1994 maxilla)                                                                                                                                                                                                                                                                                                 | <i>in situ</i>          | M. Leakey                     |                                                                                                                                                                                                                                                                                                                                                        |
| 29287                   | 1994/95 | mandible with L and RP <sub>4</sub> -M <sub>2</sub> , isolated LI <sub>1</sub> , L and RI <sub>2</sub> , RP <sub>4</sub> , L and RM <sub>3</sub>                                                                                                                                                                                                    | surface                 | S. Ngui                       | LI <sub>1</sub> : 6.8MD; LI <sub>2</sub> : 8.6MD, 8.5BL; RI <sub>2</sub> : 8.2MD; RP <sub>3</sub> : 9.2MD, >11.2 maxBL; LP <sub>4</sub> : 9.8MD, 11.4BL; RP <sub>4</sub> : 10.3BL; LM <sub>1</sub> : 12.2BLa, 12.2BLp; LM <sub>2</sub> : 13.6BLa, 12.8BLp, RM <sub>2</sub> : >14.2MD, 13.1BLa, 12.1BLp                                                 |
| 30498                   | 1995    | maxilla fragments; LI <sup>1</sup> , R <sup>C</sup> , RP <sup>3</sup> , RM <sup>1</sup> , L and RM <sub>3</sub> , L and RI, C, P, M fragments                                                                                                                                                                                                       | surface                 | K. Cheboi                     | LI <sup>1</sup> : 10.9MD, 8.2BL; R <sup>C</sup> : >10.5MD, 10.2BL; RP <sup>3</sup> : 8.7MD; 13.2BL; RM <sup>1</sup> : 11.6MD, 13.9BLa, 12.6BLp; RM <sup>3</sup> : 11.6MD, >14.3BL; LM <sup>3</sup> : 11.6MD                                                                                                                                            |
| 30500                   | 1995    | mandible fragments; L and RM <sub>1-3</sub> , LP <sub>3-4</sub> , tooth fragments                                                                                                                                                                                                                                                                   | surface                 | W. Mangao                     | LP <sub>3</sub> : 13.4maxBL; LM <sub>1</sub> : 13.3MD, 13.5BLa, 13.2BLp; RM <sub>1</sub> : 13.8MD, 13.1BLp; LM <sub>2</sub> : 15.8MD, 15.1BLa, 14.4BLp; RM <sub>2</sub> : 15.9MD, 14.7BLa, 14.2BLp; RM <sub>3</sub> : 17.0MD, 13.4BLa, 12.7BLp                                                                                                         |
| 30503                   | 1995    | proximal 3/4 manual proximal phalanx                                                                                                                                                                                                                                                                                                                | surface                 | P. Nzube                      |                                                                                                                                                                                                                                                                                                                                                        |
| 31712                   | 1996    | L and R juvenile mandible fragments; L and Rdl <sub>1-2</sub> , broken Ldm <sub>1-2</sub> , germs LI <sub>1-2</sub> , LM <sub>1</sub> , RI <sub>1</sub> , RI <sub>2</sub> , R <sub>C</sub> , RM <sub>1</sub> , tooth fragments                                                                                                                      | surface                 | W. Mangao                     | Ldm <sub>2</sub> : >10.3MD, 8.1BLp; LM <sub>1</sub> : 12.1MD, 10.5BLa, 10.2BLp; RM <sub>1</sub> : 10.5BLa, 10.0BLp; LI <sub>2</sub> : 6.6MD                                                                                                                                                                                                            |
| 31713                   | 1996    | Redentulous mandible with roots C-M <sub>2</sub> , weathered C & M crown fragments                                                                                                                                                                                                                                                                  | surface                 | P. Nzube                      |                                                                                                                                                                                                                                                                                                                                                        |
| 31714                   | 1996    | weathered Ldm <sub>2</sub>                                                                                                                                                                                                                                                                                                                          | surface                 | P. Nzube                      | Ldm <sub>2</sub> : >9.2BLa                                                                                                                                                                                                                                                                                                                             |
| 31717                   | 1996    | LM <sup>3</sup> , RM <sub>3</sub> , fragment LM <sub>2</sub>                                                                                                                                                                                                                                                                                        | surface                 | W. Mangao                     | LM <sup>3</sup> : 11.1MD, 13.8BL; RM <sub>3</sub> : 13.7MD, 12.4BLp                                                                                                                                                                                                                                                                                    |
| 31718                   | 1996    | mandible fragments with expanded and cracked M <sub>2</sub> and M <sub>3</sub>                                                                                                                                                                                                                                                                      | surface                 | L. Moru                       |                                                                                                                                                                                                                                                                                                                                                        |
| 31723                   | 1996    | worn RM <sup>3</sup>                                                                                                                                                                                                                                                                                                                                | surface                 | F. Shrenk                     | RM <sup>3</sup> : >12.7MD, >15.7BLa                                                                                                                                                                                                                                                                                                                    |
| 31724                   | 1996    | L capitate                                                                                                                                                                                                                                                                                                                                          | surface                 | P. Nzube                      |                                                                                                                                                                                                                                                                                                                                                        |
| 31726                   | 1996    | LP <sup>4</sup>                                                                                                                                                                                                                                                                                                                                     | surface                 | K. Kimeu                      | LP <sup>4</sup> : 7.2MD                                                                                                                                                                                                                                                                                                                                |
| 31727                   | 1996    | partial R <sub>C</sub> with root                                                                                                                                                                                                                                                                                                                    | surface                 | W. Mangao                     |                                                                                                                                                                                                                                                                                                                                                        |
| 31728                   | 1996    | LM <sub>1</sub>                                                                                                                                                                                                                                                                                                                                     | surface                 | W. Mangao                     |                                                                                                                                                                                                                                                                                                                                                        |
| 31729                   | 1996    | Rdm <sub>2</sub> with roots                                                                                                                                                                                                                                                                                                                         | surface                 | N. Ewalan                     | Rdm <sub>2</sub> : >9.7MD, 8.0BLa, 8.5BLp                                                                                                                                                                                                                                                                                                              |
| 31730                   | 1996    | L broken and weathered M <sub>2</sub> and RP <sub>4</sub> with roots                                                                                                                                                                                                                                                                                | surface                 | M. Muoka                      | RP <sub>4</sub> : >10BL                                                                                                                                                                                                                                                                                                                                |
| 34725                   | 1996/97 | associated juvenile dental and cranial fragments; germs L and RL <sub>1-2</sub> , L and R <sub>C</sub> , LP <sub>3</sub> , LM <sub>2</sub> , RM <sup>2</sup> , erupted crowns L and RM <sub>1</sub> , d <sub>c</sub> with roots, dm <sub>1-2</sub> , Rd <sup>2</sup> , M <sup>1</sup> , frags I <sup>1-2</sup> , P <sup>3-4</sup> , dm <sup>2</sup> | surface/ <i>in situ</i> | P. Nzube                      | permanent-I <sub>1</sub> : 7.3MD, 8.5BL; LP <sub>3</sub> : 12.1maxBL; LM <sub>1</sub> : 13.7MD, 11.9BLa, 12.2BLp; LM <sub>2</sub> : 16.0MD, 14.0BLa, 13.0BLp; RM <sup>2</sup> : 14.2MD, 14.9BLa, 14.3BLp; deciduous-Rd <sub>2</sub> : 4.8MD, 4.2BL; Ld <sub>c</sub> : 6.8MD, 5.8BL; Rd <sub>c</sub> : 7.0MD, >5.4BL; Ldm <sub>1</sub> : >9.4MD, >6.8BL |
| 35858                   | 1997    | broken RM <sub>3</sub> with roots lacking mesial and buccal portions                                                                                                                                                                                                                                                                                | surface                 | W. Mutwiwa                    |                                                                                                                                                                                                                                                                                                                                                        |
| 35839                   | 1997    | LI <sup>1</sup> with root, R <sup>C</sup> crown and partial root, broken LP <sup>3</sup>                                                                                                                                                                                                                                                            | surface/ <i>in situ</i> | P. Kapoko, W. Mangao, sievers | LI <sup>1</sup> : 12.4MD, 8.9BL; R <sup>C</sup> : 11.7MD, 11.2BL; LP <sup>3</sup> : 11.8BL                                                                                                                                                                                                                                                             |
| 35842                   | 1997    | RM <sup>1</sup> or M <sup>2</sup>                                                                                                                                                                                                                                                                                                                   | surface                 | W. Mangao                     | RM <sup>1</sup> : 11.6MD                                                                                                                                                                                                                                                                                                                               |
| 35843                   | 1997    | unerupted Ld <sub>12</sub>                                                                                                                                                                                                                                                                                                                          | surface                 | sievers                       | Ld <sub>12</sub> : 4.7MD, >4.8BL                                                                                                                                                                                                                                                                                                                       |
| 35847                   | 1997    | LM <sub>2</sub> and small tooth fragment                                                                                                                                                                                                                                                                                                            | <i>in situ</i>          | W. Mangao                     | LM <sub>2</sub> : 13.6MD                                                                                                                                                                                                                                                                                                                               |
| 35852                   | 1997    | worn L <sup>C</sup>                                                                                                                                                                                                                                                                                                                                 | surface                 | M. Eregai                     | L <sup>C</sup> : >11.0MD, >10.1BL                                                                                                                                                                                                                                                                                                                      |

Dental dimensions are standard. Deciduous dentition indicated by lower-case and permanent dentition by upper-case letters. I, incisor; C, canine; P, premolar; M, molar; L, left; R, right; MD, mesiodistal; BL, buccolingual; BLa, buccolingual anteriorly; BLp buccolingual posteriorly; max, maximum. All measurements were taken on original specimens by the authors.

comparative purposes are associated juvenile teeth and cranial pieces, a capitate, a proximal manual phalanx, further pieces of a large, presumed male mandible, and a maxilla. Among the mixed dentition of the juvenile (KNM-KP 34725) is the left lower dm1 crown with roots. The dm1 is one of the most diagnostic teeth of *Ardipithecus ramidus*<sup>3</sup>. It is diminutive and has a narrow crown with a weakly developed anterior fovea and talonid. The *A. anamensis* specimen, though worn in life, is intermediate in size and morphology between the Aramis specimen and those known at present for *A. afarensis* (Fig. 2). It is a narrow tooth with a length–breadth ratio of 1.4, close to the 1.49 ratio of *A. ramidus* and well above the mean of 1.2 for *A. afarensis*. Although wear obscures many details of the occlusal surface, the lack of distinct buccal and lingual grooves shows the talonid to have been poorly differentiated. The talonid is buccolingually expanded in *A. afarensis*.

The capitate (KNM-KP 31724) has some cortical bone missing from the dorsal surface and is broken on its palmar side. It is larger than both *A. afarensis* specimens from Hadar<sup>11</sup> and that of *A. africanus* from Sterkfontein<sup>12</sup>. It has a large, globular head and, as in all apes and *Australopithecus*, lacks the facet for the third metacarpal styloid process. It differs from the capitates of all other hominids by having a facet for the second metacarpal that faces directly laterally, as in apes. Later *Australopithecus* and *Homo* have obliquely oriented second metacarpal facets which are continuous across the dorsopalmar extent of the bone and permit some degree of rotation of the second metacarpal at this joint, an action incompatible with the morphology seen in the Kanapoi specimen (Fig. 3). The articular surface for the hamate is not obliquely set as it is in later hominids. The proximal hand phalanx (KNM-KP 30503), although incomplete and weathered, is very similar in size and morphology to those of the largest specimens in the Hadar *A. afarensis* sample. They have an equivalent degree of curvature and strong development of flexor sheath attachments.

More pieces have been collected of a large, presumed male mandible (KNM-KP 29287) which is now complete enough to

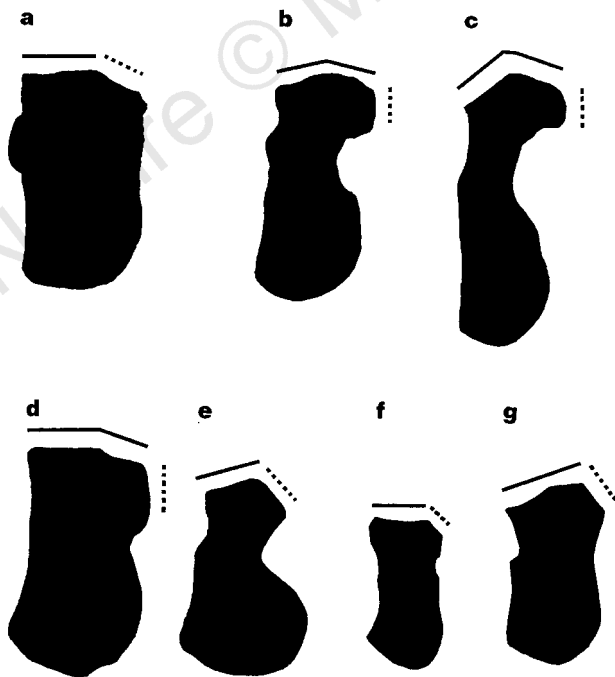
make detailed comparisons with the type, presumed female, mandible. Like the type mandible, the symphysis has its major axis set at an oblique angle, with a long postincisive plane and moderately developed superior torus. The cross-section is more similar to those of gorillas and chimpanzees than to those of later hominids. The right canine is broken with the deepest part of its root preserved. The left canine socket is empty and large (10.3 mm, mesiodistally), but a piece of its labial wall is missing, raising the possibility that it had been enlarged during fossilization. Comparisons of horizontal computed tomography (CT) scans through both bodies at the level of the mental foramen show that it is undistorted (F. Spoor, personal communication). This evidence and that from other specimens of canine crown and root indicate a large canine size dimorphism in this species. Mandibular canine crown areas are large compared with mandibular post-canine tooth crown areas. In this respect, the *A. anamensis* sample has relatively larger canines than the known *A. afarensis* sample, although the sample ranges overlap for most teeth. In addition, the range of relative canine size is similar in the two species, suggesting a comparable pattern of canine size sexual dimorphism in both. A left maxilla from Allia Bay (KNM-ER 30200) has relatively unworn teeth (the distal half of C, P<sup>3</sup>–M<sup>1</sup>, partial M<sup>2</sup>, M<sup>3</sup>) and is useful because the previously reported maxilla from Kanapoi is from an individual with very worn teeth. KNM-ER 30200 preserves the midline and confirms the primitive nature of the *A. anamensis* palate, being shallow with parallel postcanine tooth rows set close together which do not diverge posteriorly like those of all later hominids.

All hominids at Kanapoi appear to come from floodplain palaeosols or deltaic sandstones and many of them show damage caused by mammalian carnivores. Excavation after removal of the surface lag of basalt cobbles has enabled hominids to be recovered *in situ* from these palaeosols. We already have several mandibles, maxillas, skull pieces and over 50 isolated teeth of *A. anamensis* which reveal primitive character states in the skull and jaws, as well as a tibia and pieces of the forelimb (radius, distal humerus, capitate and phalanx), demonstrating that this biped retained some primitive states in its forelimb skeleton. One explanation for the hominid fossil record from nearly 4.5 Myr to just later than 3 Myr is that it represents one single, evolving species that had a gorilla-like degree of body-size sexual dimorphism. An alternative explanation is that the three recognized taxa indicate some bushiness in the early hominid tree<sup>13</sup>. Until a detailed study of *A. ramidus* is available, we cannot assess whether that taxon has any features too specialized to be ancestral to the other two. □

#### Methods

For <sup>40</sup>Ar–<sup>39</sup>Ar dating about 20 feldspar crystals for each sample were wrapped in aluminium foil and placed successively in a 6-mm-internal-diameter silica glass tube interspersed with similar packages containing fluence monitor 92-176 sanidine, separated from the Fish Canyon Tuff, Colorado, and of nominal K–Ar age 27.9 Myr<sup>14</sup>. At either end of the silica glass tube (32 mm long), a synthetic, zero-age, potassium-rich glass was included to facilitate determination of the (<sup>40</sup>Ar/<sup>39</sup>Ar)<sub>K</sub> correction factor for the particular irradiation. The silica glass tube containing the samples was encased in 0.2-mm-thick Cd shielding and placed inside an aluminium screw-topped cylinder for irradiation in the X33 or X34 position of HIFAR reactor, Lucas Heights, New South Wales. Irradiation was for 48 h to generate <sup>39</sup>Ar from <sup>39</sup>K, an essential step in the <sup>40</sup>Ar–<sup>39</sup>Ar dating technique<sup>15</sup>. At exactly 12-h intervals, the irradiation vessel was inverted end-for-end to minimize the neutron flux gradients experienced by the samples. Following irradiation, several K-glass fragments, at least five crystals of 92-176 sanidine fluence monitor from each level, and a minimum of seven alkali feldspar crystals from each sample of Kanapoi Tuff were loaded individually into wells in a copper tray, which was then installed in the ultrahigh vacuum system on-line with the VG3600 mass spectrometer.

Individual crystals of fluence monitor sanidine and alkali feldspar from the pumices in the Kanapoi Tuff, and potassium-glass fragments, were fused using a focused argon ion laser beam. The gases released were cleaned of active



**Figure 3** Coronal sections through capitates. Sections from **a**, *Homo sapiens*; **b**, *Pan troglodytes*; and **c**, *Pongo pygmaeus*, for comparison with capitates of **d**, *A. anamensis* KNM-KP 31724; **e**, *A. cf. afarensis* KNM-WT 1162; **f**, *A. afarensis* AL288-1; and **g**, *A. africanus* TM 1526, to show the shape and position of facets for metacarpal III (solid lines) and metacarpal II (dashed lines). Some figures have been reversed for ease of comparison. Scale is in mm.

constituents over SAES getters and then expanded into the VG3600 mass spectrometer, operated in the static mode, for isotopic analysis of the argon. Ion beams were detected on a Daly collector, with an overall sensitivity of the system of about  $3 \times 10^{-17}$  mol mV<sup>-1</sup>. Background in the mass spectrometer was negligible. At least five measurements were made on the fluence monitor at each level, with concordant results shown by uncertainties of <0.4% standard deviation of the <sup>40</sup>Ar/<sup>39</sup>Ar<sub>K</sub> ratios, the ratio of radiogenic argon to K-derived <sup>39</sup>Ar. The fluence gradient measured over the sample interval in the silica glass tube is about 1.4%, and individual interpolated uncertainties for the fluence at each level was <0.35%, used in the calculations.

Received 13 November 1997; accepted 2 February 1998.

1. Leakey, M. G., Feibel, C. S., McDougall, I. & Walker, A. New four million-year-old hominid species from Kanapoi and Allia Bay, Kenya. *Nature* **376**, 565–571 (1995).
2. Andrews, P. Paleontology—Ecological apes and ancestors. *Nature* **376**, 555–556 (1995).
3. White, T. D., Suwa, G. & Asfaw, B. *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature* **371**, 306–312 (1994).
4. White, T. D., Suwa, G. & Asfaw, B. Corrigendum: *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature* **375**, 88 (1995).
5. Johanson, D. C. & White, T. D. A systematic assessment of early African hominids. *Science* **202**, 321–330 (1979).
6. Kimbel, W. H., Johanson, D. C. & Rak, Y. The first skull and other new discoveries of *Australopithecus afarensis* at Hadar, Ethiopia. *Nature* **368**, 449–451 (1994).
7. Johanson, D. C., White, T. D. & Coppens, Y. Dental remains from the Hadar Formation, Ethiopia: 1974–1977 collections. *Am. J. phys. Anthropol.* **57**, 545–603 (1982).
8. Brown, F. H. & Feibel, C. S. Stratigraphy, depositional environments and palaeogeography of the Koobi Fora Formation. In *Koobi Fora Research Project Vol. 3* (ed. Harris, J. M.) 1–30 (Clarendon, Oxford, 1991).
9. Powers, D. W. *Geology of Miocene Sediments of the Lower Kerio Valley, Kenya* (Thesis, Univ. Princeton, 1980).
10. Haileab, B. *Geochemistry, Geochronology and Tephrostratigraphy of Tephra from the Turkana Basin, Southern Ethiopia and Northern Kenya* (Thesis, Univ. Utah, 1995).
11. Bush, M. E., Lovejoy, C. O., Johanson, D. C. & Coppens, Y. Hominid carpal, metacarpal, and phalangeal bones recovered from the Hadar Formation: 1974–1977 collections. *Am. J. phys. Anthropol.* **57**, 651–677 (1982).
12. Robinson, J. T. *Early Hominid Posture and Locomotion* (University of Chicago Press, Chicago, 1972).
13. Gould, S. J. Unusual unity. *Nat. Hist.* **106**, 20–23 and 69–70 (1997).
14. Cebula, G. T. et al. The Fish Canyon Tuff, a potential standard for the <sup>40</sup>Ar/<sup>39</sup>Ar and fission track dating methods. (Abstract) *Terra Cognita* **6**, 139–140 (1986).
15. McDougall, I. & Harrison, T. M. *Geochronology and Thermochronology by the <sup>40</sup>Ar/<sup>39</sup>Ar Method* (Oxford Univ. Press, New York, 1988).

**Acknowledgements.** We thank the Government of Kenya for permission to carry out this research and the National Museum of Kenya for logistical support. The National Geographic Society funded the field work; the National Science Foundation and the Leakey Foundation awarded grants to C.F. and A.W. Neutron irradiations were facilitated by the Australian Institute of Nuclear Science and Engineering and the Australian Nuclear Science and Technology Organization. We thank Caltex (Kenya) for fuel and Rover Group and the NGS for vehicles, and T. White and F. Spoor for their help. The expedition included K. Cheboi, S. Dominic, G. Ekalale, M. Erege, J. Erus, N. Ewalan, J. Kaatho, C. Kiarie, K. Kimeu, F. Kyalo, M. Kyeva, B. Kyongo, L. Leakey, N. Lewis, S. Makathimo, B. Malika, W. Mangao, S. Milledge, L. Moru, M. Muoka, J. Mutaba, D. Mutinda, N. Mutiwa, S. Ngui, B. Onyango and J. Wynn; S. Hagemann and L. Hlusko assisted at Allia Bay; J. Mya undertook the mineral separations; R. Maier assisted with handling of samples for irradiation; and A. Ibui, F. Kyalo, E. Mbua, M. Muungu and A. Mwai provided curatorial assistance.

Correspondence and requests for materials should be addressed to M.G.L. (e-mail: mifupa@iconnect.co.ke).

## Probing motivational state during agonistic encounters in animals

R. W. Elwood, K. E. Wood, M. B. Gallagher & J. T. A. Dick

School of Biology and Biochemistry, The Queen's University of Belfast, Belfast BT9 7BL, UK

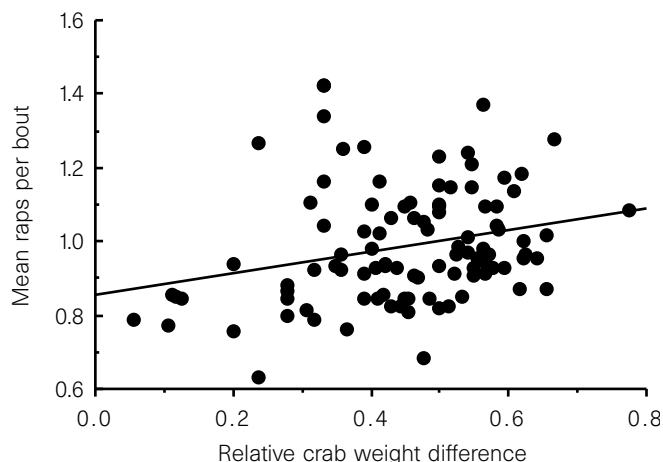
Animals commonly compete for resources by direct aggression: for example, spiders fight for web sites<sup>1</sup>, male red deer fight for females<sup>2</sup>, and scorpionflies fight for prey<sup>3</sup>. The application of game theory has considerably advanced our understanding of the evolution of such contests<sup>4–7</sup>. A general conclusion is that, if possible, animals should assess both the relative fighting abilities and the value of resources before making tactical decisions during contests<sup>8</sup>. These tactical decisions are assumed to be mediated by differing motivational state<sup>7,9</sup>, but this fundamental assumption has yet to be tested. Here we test the accumulated theory by

probing the motivational state of hermit crabs during fights over the ownership of gastropod shells. The test uses a stimulus, novel to the crabs, that produces a startle response, the duration of which is an independent measure of the motivation to fight. We demonstrate that motivational state differs at an early stage of the contest according to the potential gain in resource value. There was no effect of relative size of the opponent on motivational state. In these contests, relative size neither predicted the likely cost of the contest nor the probability of victory.

During the course of any single activity, the motivational state of an animal may vary<sup>10</sup>, so that subsequent behavioural change is a retrospective indication of motivational change. Thus, little information about motivational state may be available from ongoing behaviour<sup>11</sup>. In fights, it may be important not to signal future intentions<sup>12</sup>, so that behaviour is even less likely to reflect motivational state. Therefore, to assess the motivational state of fighting animals, we need a measure that is independent of the ongoing behaviour. This may be achieved by the introduction of a potentially startling stimulus, which has been used before to investigate motivational state during feeding and preening in chicks<sup>10</sup>, walking in locusts<sup>13</sup> and assessment of empty shells in hermit crabs<sup>14</sup>. These studies indicated that the duration and/or severity of the startle response is inversely related to the motivation of the animal to continue the previous activity. Our investigation uses this validated measure to probe the motivational state of animals during the course of a fight. We determine whether relative fighting ability (relative size) and resource value are assessed by the animal and, if so, how the information influences motivational state at a particular point in the contest.

European hermit crabs (*Pagurus bernhardus*), of the size used in this study, show a marked preference for periwinkles (*Littorina obtusata*) over top shells (*Gibbula cineraria*). They also show preferences for a particular size of shell relative to crab size<sup>15,16</sup> and crabs in preferred shells have a high reproductive success<sup>17,18</sup>. Most suitable shells are occupied by crabs, and fights for ownership are common. In a typical fight, the larger crab (attacker) approaches and grasps the shell of the smaller crab (defender), causing the defender to withdraw into its shell. The attacker feels over the exterior of the defender's shell, turns the shell and feels into the aperture. The attacker may escalate the contest by repeatedly rapping the two shells together in discrete bouts, interspersed with pulling at the defender's chelipeds. This may cause the defender to release its abdominal grip on the shell and it is then pulled out by the attacker, allowing the latter to take the defender's shell. Alternatively, the attacker may give up the fight at any stage.

The absolute quality of the attacker's shell, but not the defender's shell, influences the decision to attack, but the subsequent decision



**Figure 1** The mean number of raps per bout (log) increases as the relative crab weight difference (log) increases ( $y = 0.298x + 0.855$ ;  $r^2 = 0.067$ ;  $P < 0.01$ ).



to escalate depends upon the potential gain in shell quality<sup>19</sup>. This suggests that the attacker assesses the quality of the defender's shell relative to that of its own after the start of the fight but before escalation. The relative size of the crabs also influences tactical decisions in fights<sup>20,21</sup>, indicating that the relative resource holding potential<sup>8</sup> must be assessed. Hermit crab fights, in common with theory<sup>6-8</sup>, may thus be viewed as a process of information gathering with consequent effects on motivational state. We now study the motivational state of attacking crabs before they decide whether or not to escalate the fight.

Crabs were collected from a shore in Northern Ireland and removed from their original *L. obtusata* shells and each was given either a *G. cineraria* or a *L. obtusata* shell to occupy. Only male crabs were allocated in pairs of different-size animals to one of two experimental groups. In all cases, the smaller crab of the pair was given a *L. obtusata* shell of the correct size for the larger crab. In group A ( $n = 92$ ), the larger crab of each pair was given a *G. cineraria* shell that was too small for it, although the shell was the correct size for the smaller crab. In group B ( $n = 117$ ), the larger crab was given a *G. cineraria* shell that was the correct size for itself and too large for the smaller crab. Thus, in each case, the larger crab could obtain the preferred size and species by taking the opponent's shell, but the gain would be greater in group A than in group B. Preferred shells were determined from regression analysis of previous shell-selection experiments<sup>21</sup>.

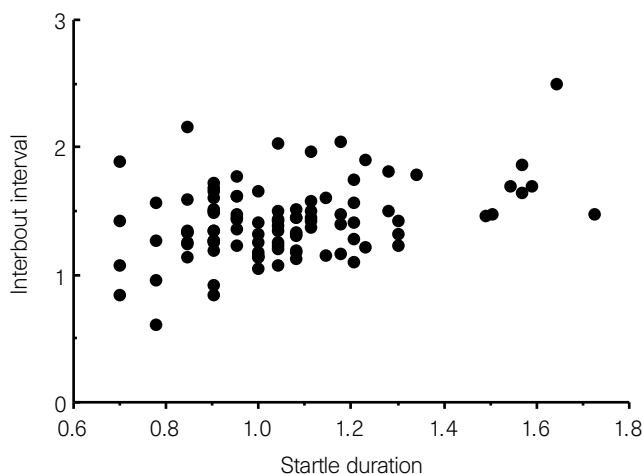
Crabs were kept in isolation for about 18 hours before being placed with their opponent in a small glass dish containing sea water and a sand substrate. The activities of the crabs were recorded on a Psion computer using the Observer package (Noldus Technology). If no fight occurred within 30 minutes, the observation was terminated; however, if a fight started and the attacker had felt into the aperture of the defender's shell for 10 seconds, the stimulus was applied. This was a weighted black card (12.5 cm<sup>2</sup>; 83 g) dropped from a height of 34.5 cm onto the side of the glass dish. This consistently caused the attacker temporarily to stop fighting and withdraw into its shell before resuming the fight.

The two groups did not differ in the probability of a fight (A: 59/92 (64%); B: 62/117 (53%); G test, NS, power, 0.36) but, of those that fought, there were significantly more evictions in the high-gain group (A: 53/59 (90%); B: 47/62 (76%); G test, 4.27;  $P < 0.05$ ). Analysis of covariance was used to examine the effects of potential gain for the attacker (a discrete factor) and of the relative size of the opponents (the regressor) on the startle duration of the attacker for all fights, and on various parameters of shell rapping for those that resumed the fight and effected an eviction (all behavioural

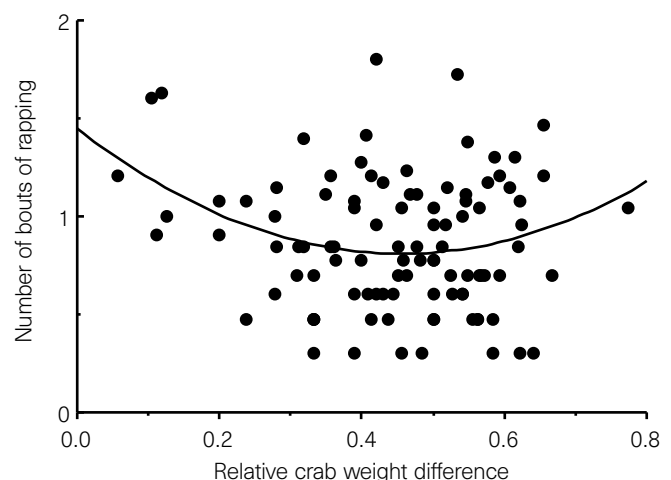
data were log<sub>10</sub>-transformed for analysis). Group A (high gain) crabs had significantly shorter startle responses than group B (low gain) ( $F_{1,118} = 7.53$ ;  $P < 0.01$ ) (untransformed means, 10.17 s and 14.26 s, respectively), but there was no effect of relative size of crabs (determined as 1 minus the weight of the small crab divided by the weight of the large crab). Crabs in group A fought harder than those in group B, as measured by the total number of raps required for eviction (untransformed means were 93.038 and 59.064, respectively) ( $F_{1,97} = 6.13$ ;  $P < 0.02$ ) and the number of bouts of shell rapping (untransformed means were 11.151 and 7.106, respectively) ( $F_{1,97} = 7.19$ ;  $P < 0.01$ ) but not by the number of raps per bout. There was no effect of relative crab size on the total raps or on the number of bouts of rapping required to effect an eviction, although fewer raps per bout occurred with more evenly matched crabs ( $F_{1,97} = 7.25$ ;  $P < 0.01$ ) (Fig. 1).

The shorter startle response in the high-gain group indicates that information about potential gain is available to attacking crabs before escalation of the contest. The motivational state thus differs according to potential gain, as measured by motivation that is independent of behaviour during contest. Furthermore, the high level of motivation in group A appears to persist throughout the contest, as indicated by the greater effort invested. Group A crabs gave more raps in total and more bouts of rapping. Surprisingly, there was no difference between the groups in the duration of the contest, measured from the end of the startle response to eviction of the defender. Rather, Group A crabs fought more vigorously than those in group B, with shorter gaps between bouts of rapping (untransformed means were 25.592 s and 41.269 s, respectively) (ANCOVA  $F_{1,97} = 3.916$ ;  $P = 0.05$ ), resulting in more raps in the same time. Furthermore, the duration of the startle was positively correlated with the average interval between bouts. This was found overall ( $r = 0.37$ ; d.f. = 98;  $P < 0.01$ ) (Fig. 2) and also for each group ( $r = 0.415$ ; d.f. = 51;  $P < 0.01$ , and  $r = 0.316$ ; d.f. = 45;  $P < 0.05$  for A and B, respectively). This suggests that startle duration and vigour of the fight both reflect the motivation to obtain the shell.

In apparent contradiction with theory<sup>7</sup>, the relative size of the combatants did not appear to influence the motivation of the attacker. Relative size is assessed before a fight, however, because virtually all fights are initiated by the larger crab (less than 5% of observations were excluded from the present analysis because the smaller crab initiated the fight). Furthermore, the temporal pattern of the fight varied with relative size, with fewer raps per bout given to relatively large opponents (Fig. 1), although there was no effect on other measures. In particular, relative size failed to influence total



**Figure 2** The mean interbout interval (log) is positively correlated with the startle duration (log) ( $r = 0.37$ ; d.f. = 98;  $P < 0.01$ ).



**Figure 3** The number of bouts of shell rapping (log) shows a second-order polynomial relationship with relative crab weight difference ( $y = 1.447 - 2.805x + 3.085x^2$ ;  $r^2 = 0.072$ ;  $P = 0.05$ ).

effort to gain an eviction and was not a predictor of victory. This result differs from those obtained in other animal contests but is easily explained. The smaller crab was always housed in a shell of the right size for the larger crab, so when there was a large size difference between the crabs the small crab was in a shell that was much too large for it. This enabled the smaller crab to withdraw deeply into the shell, making it difficult for the attacker to pull at it during the contest. A second-order polynomial regression between relative size and the number of bouts of rapping required for eviction makes this point. This shows a clear U-shaped function ( $F_{2,97} = 3.757$ ;  $r^2 = 0.072$ ;  $P < 0.05$ ) (Fig. 3), with particularly large or small size differences resulting in more bouts of shell rapping. Thus, relative crab size does not appear to affect motivational state, at least at this stage of the contest, perhaps because it neither readily predicts the potential cost of the contest nor the probability of victory.

Hermit-crab fights have been interpreted as being a process of negotiation rather than aggression because both crabs could stand to gain from a shell exchange<sup>22,23</sup>. This might happen if a large crab in a shell that was too small were to fight a small crab in a shell that was too large. It has been shown for other species of hermit crab that the probability of an exchange increases if there is a potential gain for the defender (or non-initiator)<sup>23</sup>. In our study, the defenders that exchanged in group A obtained a shell of the right size, but in group B they obtained one that was too large. The negotiation model thus predicts the greater number of exchanges seen in group A; however, it also predicts that defenders should have been more willing to exchange in group A and should have resisted for a shorter time, which did not occur—durations were the same but attackers fought more vigorously in group A. Our findings are therefore in agreement with the view that shell fights are a process of aggression<sup>21</sup>. It seems that it is the high motivation of group A attackers that leads to a greater effort in fighting, resulting in a higher probability of victory. There is no evidence in this species that defenders can obtain information about the attacker's shell and modify their behaviour accordingly.

The startle technique enables the motivation of participants to be measured at an early stage of the fight and the resulting startle duration correlates with the subsequent vigour of the fight. Furthermore, the startle response agrees with game theory in that motivational state reflects information about resource value<sup>7</sup>. The technique, however, produced a surprising result in that relative crab size did not appear to influence motivation, but that is likely to be peculiar to hermit-crab fights for reasons already outlined. The technique could be applied to contests in other species, in which relative size is the key determinant of persistence of each combatant. It could also be used to show how motivation changes as information about the opponent's fighting ability is gathered. For example, it is assumed in the theory that information about relative fighting ability improves with increased time spent in the fight<sup>7</sup>. If both participants are startled, the prediction is that the smaller animal should have the longer startle duration and that, the further into the contest, the greater should be the disparity in startle duration between two contestants. Thus, by varying the timing of the novel stimulus, we may determine at which stage in the contest animals obtain information and how that information shapes subsequent behaviour. Hence, this technique allows basic assumptions of game theory, as applied to animal conflicts, to be tested. □

Received 19 December 1997; accepted 6 February 1998.

- Riechert, S. E. Games spiders play. II. Resource assessment strategies. *Behav. Ecol. Sociobiol.* **6**, 121–125 (1979).
- Clutton-Brock, T. H., Albon, S. D., Gibson, R. M. & Guinness, F. E. The logical stag: Adaptive aspects of fighting in the red deer (*Cervus elaphus* L.). *Anim. Behav.* **27**, 211–225 (1979).
- Thornhill, R. Fighting and assessment in *Harpo bitticus* scorpionflies. *Evolution* **38**, 204–214 (1984).
- Maynard Smith, J. & Price, G. R. The logic of animal conflict. *Nature* **246**, 15–18 (1973).
- Maynard Smith, J. The theory of games and the evolution of animal conflict. *J. Theor. Biol.* **47**, 209–221 (1974).
- Parker, G. A. & Rubenstein, D. I. Role assessment, reserve strategy and acquisition of information in asymmetric animal contests. *Anim. Behav.* **29**, 221–240 (1981).

- Enquist, M. & Leimar, O. Evolution of fighting behaviour: decision rules and assessment of relative strength. *J. Theor. Biol.* **102**, 387–410 (1983).
- Parker, G. A. Assessment strategy and the evolution of fighting behaviour. *J. Theor. Biol.* **47**, 223–243 (1974).
- Parker, G. & Stuart, R. A. Animal behaviour as a strategy optimiser: evolution of resource assessment strategies and optimal emigration thresholds. *Am. Nat.* **110**, 1055–1076 (1976).
- Culshaw, A. D. & Broom, D. M. The imminence of behavioural change and the startle response of chicks. *Behaviour* **73**, 64–76 (1980).
- Sibly, R. & McCleery, R. H. The dominance boundary method of determining motivational state. *Anim. Behav.* **24**, 108–124 (1976).
- Maynard Smith, J. & Parker, G. A. The logic of asymmetric contests. *Anim. Behav.* **24**, 159–175 (1974).
- Moorehouse, J. E., Fosbrooke, I. H. M. & Ludlow, A. R. Stopping a walking locust with sound: an analysis of variation in behavioural threshold. *Exp. Biol.* **46**, 193–201 (1987).
- Jackson, N. W. & Elwood, R. W. Interrupting an assessment process to probe changes in motivational state. *Anim. Behav.* **39**, 1068–1077 (1990).
- Elwood, R. W., McClean, A. & Webb, L. The development of shell preferences by the hermit crab *Pagurus bernhardus*. *Anim. Behav.* **27**, 940–946 (1978).
- Elwood, R. W. Motivational change during resource assessment by hermit crabs. *J. Exp. Mar. Biol. Ecol.* **193**, 41–55 (1995).
- Hazlett, B. A. & Baron, L. C. Influence of shells on mating behaviour in the hermit crab *Calcinus tibicen*. *Behav. Ecol. Sociobiol.* **24**, 369–376 (1989).
- Elwood, R. W., Marks, N. & Dick, J. T. A. Consequences of shell-species preferences for female reproductive success in the hermit crab *Pagurus bernhardus*. *Mar. Biol.* **123**, 431–434 (1995).
- Dowds, B. M. & Elwood, R. W. Shell wars: assessment strategies and the timing of decisions in hermit crab shell fights. *Behaviour* **85**, 1–24 (1983).
- Elwood, R. W. & Glass, C. W. Negotiation or aggression during shell fights of the hermit crab *Pagurus bernhardus*? *Anim. Behav.* **29**, 1239–1244 (1981).
- Elwood, R. W. & Neil, S. J. *Assessments and Decisions: A Study of Information Gathering by Hermit Crabs* (Chapman & Hall, London, 1992).
- Hazlett, B. A. Information transfer during shell exchange in the hermit crab *Clibanarius antillensis*. *Anim. Behav.* **35**, 218–226 (1987).
- Hazlett, B. A. Assessments during shell exchanges by the hermit crab *Clibanarius vittatus*: the complete negotiator. *Anim. Behav.* **51**, 567–573 (1996).

**Acknowledgements.** We are grateful to the Association for the Study of Animal Behaviour for funding this work.

Correspondence and requests for materials should be addressed to R.W.E. (e-mail: r.elwood@qub.ac.uk).

## Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid

Rachel J. Waugh O'Neill\*†, Michael J. O'Neill‡ & Jennifer A. Marshall Graves\*

\* Department of Genetics and Human Variation, La Trobe University, Bundoora, Victoria 3083, Australia

‡ Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

Genetic models<sup>1,2</sup> predict that genomic rearrangement in hybrids can facilitate reproductive isolation and the formation of new species by preventing gene flow between the parent species and hybrid (sunflowers are an example<sup>3</sup>). The mechanism underlying hybridization-induced chromosome remodelling is as yet unknown, although mobile element activity has been shown to be involved in DNA rearrangement in some dysgenic *Drosophila* hybrids<sup>4,5</sup>. It has been proposed that DNA methylation evolved as a means of repressing the movement of mobile elements (the host defence model<sup>6,7</sup>). If such a protective mechanism were to fail, mobile elements could be activated, and could cause major and rapid genome alterations<sup>8,9</sup>. Here we demonstrate the occurrence of genome-wide undermethylation, retroviral element amplification and chromosome remodelling in an interspecific mammalian hybrid (*Macropus eugenii* × *Wallabia bicolor*). Atypically extended centromeres of *Macropus eugenii* derived autosomes in the hybrid were composed primarily of an unmethylated, amplified retro-

† Present address: Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA.

viral element not detectable in either parent species. These results, taken with the observation of deficient methylation and *de novo* chromosome change in other mammalian hybrids, indicate that the failure of DNA methylation and subsequent mobile-element activity in hybrids could facilitate rapid karyotypic evolution.

Marsupials, with their few, large chromosomes, are ideal for investigation of the incidence and cause of genome rearrangement in interspecific hybrids. Over the past 20 years, there have been several reports of hybrids between macropodid (kangaroo) species with different karyotypes. In many of these hybrids, *de novo* chromosome abnormalities are apparent even without modern staining/banding methods. For instance, three interspecies hybrids (*Macropus agilis* × *M. rufogriseus*, *W. bicolor* × *M. rufogriseus*, *M. eugenii* × *M. dorsalis*) exhibited striking centromere expansions on the chromosomes of one parental set<sup>10</sup>. We now describe the cytogenetic and molecular analysis of the hybrid male, BE-1, produced from a mating of two kangaroo species, *Wallabia bicolor* (swamp wallaby) and *Macropus eugenii* (tammar wallaby). We sought evidence of a correlation between hybridization-associated genome reorganization, mobile-element activity and DNA methylation. The hybrid BE-1 is a sterile male with no gross abnormalities other than an absence of spermatogenesis (M. Eldridge, personal communication). As the parent animals were no longer available, several individuals from the same populations as the parents were used in comparisons with the hybrid.

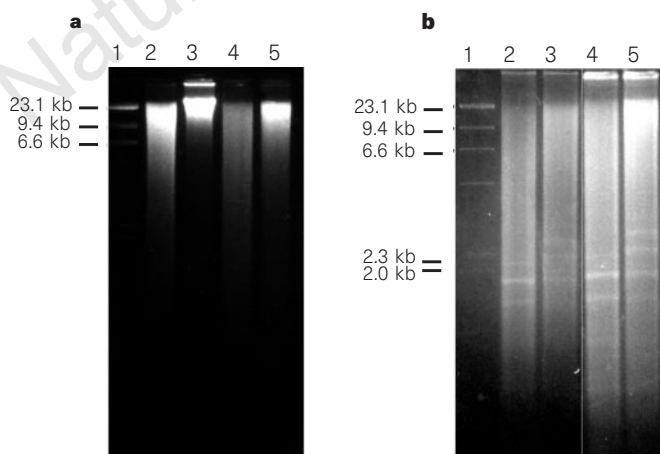
In order to detect and characterize *de novo* chromosome rearrangements in this species hybrid, DAPI (4,6-diamidino-2-phenylindole)-stained chromosome preparations from whole-blood cultures of the hybrid were compared with those of the parent species. Chromosomes of the tammar wallaby (2N = 16) and swamp wallaby (2N = 10M, 11F) can all be readily distinguished in the hybrid by size and centromere position. All the swamp-wallaby-derived chromosomes in the hybrid appeared karyotypically normal<sup>11</sup>. However, all the chromosomes of tammar origin in the hybrid had strikingly extended centromeres compared with those of the tammar wallaby. In over 25 years of cytological studies of the tammar wallaby, no such centromeric extension have been observed. The extensions seen in this hybrid are therefore *de novo* rearrangements specific to the hybrid, resulting from amplification of sequences juxtaposed or integral to the centromere. Other *de novo* rearrangements were detected by *in situ* hybridization

with isolated tammar wallaby chromosome paints<sup>11</sup>, including a chromosome 2 to chromosome 7 transposition and rearrangement of Xp across the NOR region (R.J.W.O'N. *et al.*, manuscript in preparation).

The methylation status of parental and hybrid genomic DNA, isolated from both whole blood and fibroblasts, was compared by digestion with the restriction enzymes *MspI* and *HpaII*. Both enzymes cut at CCGG, but *HpaII* does not cut if the site is methylated. Electrophoresis of tammar and swamp wallaby DNA showed digestion patterns typical of mammalian genomic DNA. A smear in the *MspI* lane denoted complete digestion, whereas mostly uncut DNA at the top of the *HpaII* lane indicated extensive methylation of CpGs (Fig. 1a). In striking contrast, the digestion of hybrid DNA with either enzyme resulted in a smear from the well to the DNA front and a ladder of satellite bands (Fig. 1b). The extensive cutting of hybrid DNA by *HpaII* indicated a significant loss of methylcytosine, and the satellite bands suggested that unmethylated and highly amplified sequences were present in the hybrid but absent from both parent species.

To isolate these unmethylated, amplified sequences, satellite bands were excised from the *HpaII*-digested hybrid genomic DNA and subcloned (Fig. 1b). Double-stranded sequence was obtained for a 2-kilobase (kb) band, and for several bands in the size range 200–400 base pairs (bp). A search of the Genbank database with the nucleotide sequence of the 2-kb clone yielded greatest similarity to the *pol* regions of a human endogenous retrovirus, HERV-K10, and the Y-chromosome-linked mouse interstitial A particle, the MIAP-Y retrovirus. A search of protein databases using the deduced amino-acid sequence of the 2-kb clone yielded significant homology to the *gag*, *prt* and/or *pol* genes of many retroviruses, representing many classes and types of the MMTV family. This suggests that the amplified sequence was derived from a fragment of a retrovirus, or retrotransposon, containing parts of the *gag*, *prt*, and *pol* genes. This clone represents a novel retroviral element, which we have named KERV-1 (kangaroo endogenous retroviral element-1). Each of the smaller clones represented a derivative fragment of the larger, 2-kb clone, with ~92% similarity. The slight sequence variation of these fragments is typical of the rapid evolution of RNA genomes<sup>12,13</sup> and reflects the extremely high level of amplification this element has undergone.

The 2-kb clone was used to probe Southern blots of digested DNA



**Figure 1** Restriction/methylation analysis of genomic DNA. **a**, Digestion of *Macropus eugenii* (tammar wallaby) and *Wallabia bicolor* (swamp wallaby) genomic DNA with *MspI* and *HpaII*. Lane 1,  $\lambda$ -HindIII markers; lanes 2 and 3, tammar wallaby (animal number 3221) genomic DNA digested with *MspI* and *HpaII*; lanes 4 and 5, swamp wallaby (110) genomic DNA digested with *MspI* and *HpaII*. **b**, Digestion of hybrid genomic DNA isolated from either whole blood (lanes 2 and 3) or fibroblasts (lanes 4 and 5) with *MspI* (lanes 2 and 4) or *HpaII* (lanes 3 and 5). Lane 1,  $\lambda$ -HindIII markers.



**Figure 2** Southern analysis of the element cloned from hybrid genomic DNA to genomic DNA from the hybrid digested with *MspI* (lane 1) and *HpaII* (lane 2). Markers are listed on the left. Autoradiography was at room temperature for 15 min.

from the parent species and the hybrid. KERV-1 hybridized strongly to satellite bands of *MspI*- and *HpaII*-digested hybrid DNA (Fig. 2), confirming that it represented the original excised band and contained DNA that was highly repeated and unmethylated in the hybrid genome. The hybridization of this probe to many bands again suggests that this element has undergone frequent mutation during reverse transcription and amplification. Not all satellite bands present in the original digests hybridized to this clone, indicating the presence of other unmethylated and amplified elements in the hybrid.

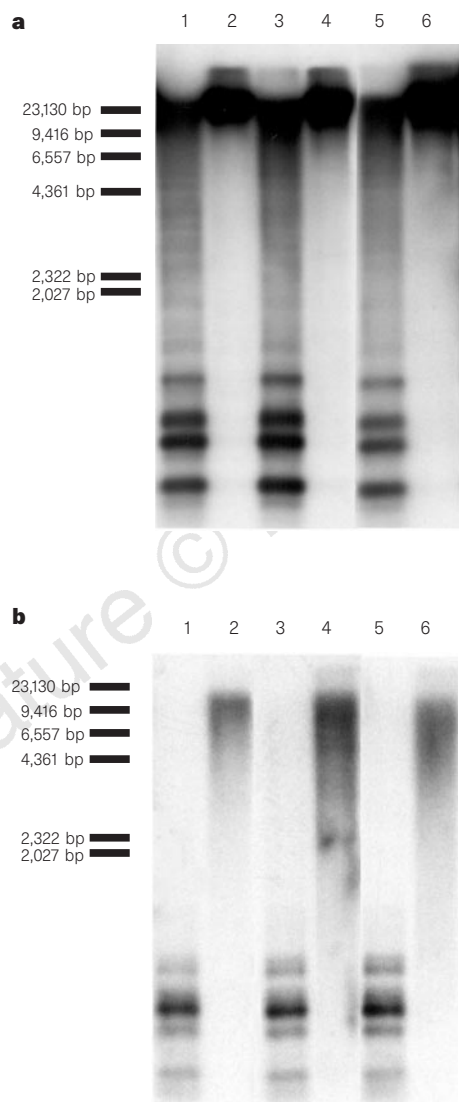
To identify KERV-1 within the parental genomes, genomic DNA isolated from whole blood from seven tammar wallabies representing two distantly related populations (Garden Island and Kangaroo Island), and from six swamp wallabies (unrelated individuals from a captive population), was digested with *MspI*, *HpaII*, *BamHI*, and *PstI*. No individual from either parent species shared *BamHI*, *PstI*

bands with the hybrid (data not shown). Hybridization of KERV-1 to *MspI*- but not *HpaII*-digested DNA revealed small fragments in the tammar and swamp wallaby genomes (Fig. 3a, b). This suggests that the bands contain a methylated source element that has undergone sequence changes in the hybrid, resulting in different restriction-enzyme digestion patterns. Alternatively, the amplified element in the hybrid could be a product of xenologous recombination between two partial viruses<sup>14</sup> present in one or both parents. Less likely is the possibility that the element is the result of horizontal transfer and subsequent integration.

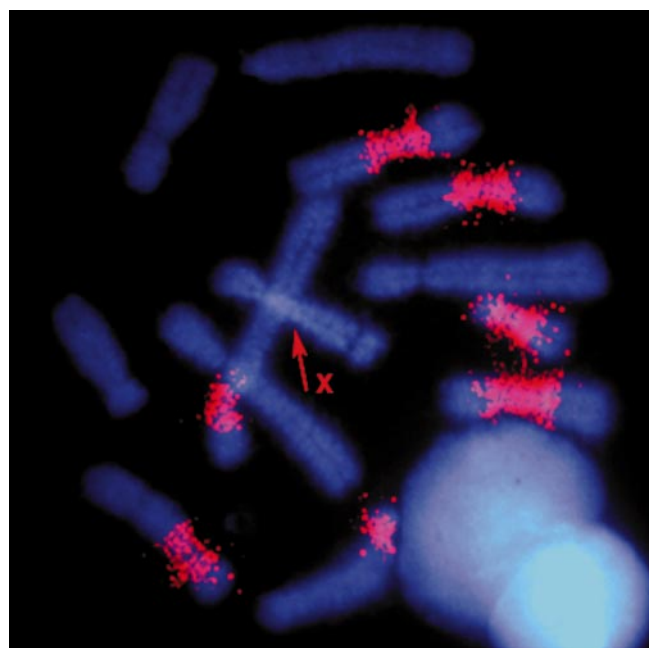
An obvious candidate for the amplified centromeric element observed in the DAPI-stained chromosomes of the hybrid was the unmethylated and amplified sequence, KERV-1, cloned from the digests of hybrid genomic DNA. Fluorescence *in situ* hybridization using KERV-1 as a probe for hybrid chromosomes confirmed the concentration of highly repeated KERV-1 in the extended centromeres of all the tammar-derived autosomes (Fig. 4). This element was not present in the extended centromere of the tammar X chromosome, which must therefore contain a second, as-yet unidentified, element.

In contrast, *in situ* hybridization of KERV-1 to chromosomes of the two parent species showed no consistent signals under high- or low-stringency hybridization conditions. Weak and variable hybridization to a few swamp-wallaby-chromosome arms, observed only under low-stringency conditions, may denote a related element or nonspecific binding. KERV-1 showed no detectable hybridization signal on any of the tammar wallaby chromosomes under either low- or high-stringency conditions. Thus, no sequence homologous to that of the retroviral element isolated from the hybrid genome could be detected in either parent species (data not shown). This again suggests that the KERV-1 element was produced by xenologous recombination and was transposed into the centromeres of the tammar-wallaby-derived autosomes of this hybrid. Retroviral integration into only the tammar centromeres could have been produced by interaction of retroviral integrase with species-specific heterochromatin<sup>15,16</sup>.

The appearance of bands in both the *MspI* and the *HpaII* digests



**Figure 3** Examples of Southern analysis of KERV-1 to genomic DNA digested with *MspI* and *HpaII* from several tammar and swamp wallabies. Both autoradiographs were exposed to film overnight with an intensifying screen at  $-70^{\circ}\text{C}$ . Markers are listed on the left. **a**, Lanes 1 and 2, *W. bicolor* (Wb2) DNA digested with *MspI* and *HpaII*; lanes 3 and 4, *W. bicolor* (Wb3) DNA digested with *MspI* and *HpaII*; lanes 5 and 6, *W. bicolor* (Wb5) DNA digested with *MspI* and *HpaII*. **b**, Lanes 1 and 2, *M. eugenii* (3206) DNA digested with *MspI* and *HpaII*; lanes 3 and 4, *M. eugenii* (T151) DNA digested with *MspI* and *HpaII*; lanes 5 and 6, *M. eugenii* (3295) DNA digested with *MspI* and *HpaII*.

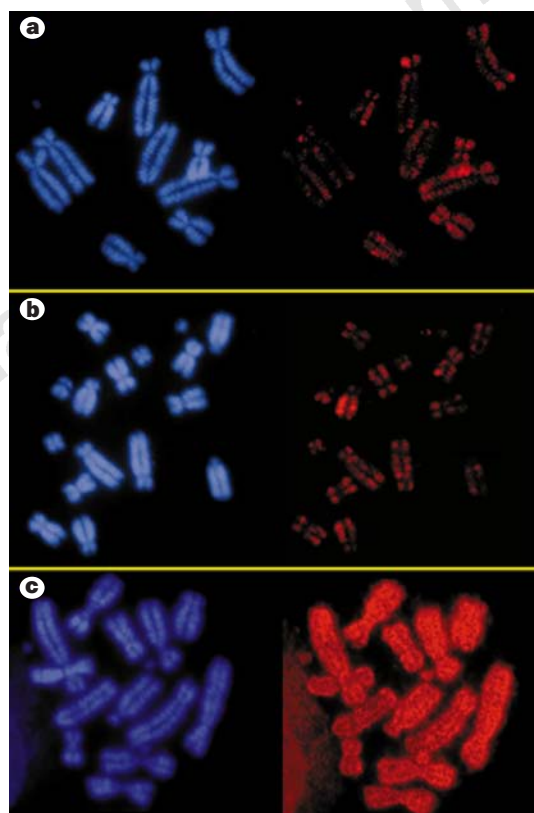


**Figure 4** Fluorescence *in situ* hybridization of KERV-1 obtained from genomic DNA of the hybrid. Hybridization of digoxigenin-labelled probe to the centromeres of all of the tammar-wallaby-derived autosomes. The unlabelled tammar X chromosome is indicated by a red arrow.

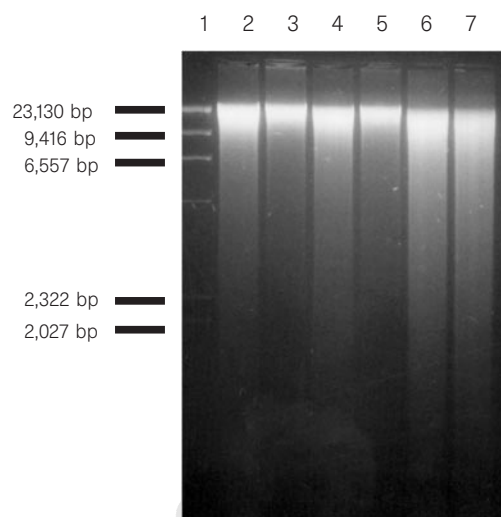


of the hybrid DNA indicates that amplified elements, including KERV-1, were unmethylated. However, a more extensive under-methylation of the hybrid genome compared to the parent species was suggested by the background smear of the *Hpa*II digest. We adapted *in situ* nick translation<sup>17</sup> using a methyl-sensitive enzyme to study the distribution of unmethylated moieties (*Hpa*II sites) in the genome. After *Hpa*II digestion of chromosomes on a slide, digoxigenin was incorporated into the cut sites with DNA polymerase I and detected with FAB (anti-digoxigenin–rhodamine) fragments. As digoxigenin is incorporated only into the unmethylated CpG sites at which *Hpa*II cuts, unmethylated regions appear red whereas methylated regions remain unlabelled.

*In situ* nick-translations were performed on chromosome preparations from a male swamp wallaby, a male tammar wallaby, and the interspecific hybrid, with human chromosomes used as a control. The signals obtained for tammar wallaby and swamp wallaby had the same moderate intensity and punctate distribution as the human control (see Fig. 1 in Supplementary Information), and represented a baseline fluorescence signal typical of a wide range of mammals (Fig. 5a, b). Under identical conditions, the signal over chromosomes of the interspecific hybrid was much more intense than in the parental and human controls (Fig. 5c), and was equivalent to signal obtained by *Msp*I digestion of parental chromosomes (data not shown). The uniformly high level of incorporation indicated an extremely low level of methylation throughout the hybrid genome. Undermethylation was not confined to the tammar centromeres, which contained the amplified mobile elements, but extended the length of all chromosomes derived from both parents. Although this method may not detect small, unlabelled (methylated) regions within regions of intense fluorescence, it was ade-



**Figure 5** *In situ* nick-translation on metaphase chromosomes from wallabies. **a**, Swamp wallaby; **b**, tammar wallaby; and **c**, the hybrid male. DAPI (4,6-diamidino-2-phenylindole)-stained chromosomes are shown on the left; the same spread after digoxigenin labelling and detection with anti-digoxigenin–rhodamine fragments on chromosomes digested with *Hpa*II are on the right.



**Figure 6** Digests of genomic DNA from two rock wallaby species and the  $F_1$  hybrid between them. Lane 1,  $\lambda$  *Hind*III marker; lanes 2 and 3, *Petrogale godmani* DNA digested with *Msp*I and *Hpa*II; lanes 4 and 5, *P. mareeba* DNA digested with *Msp*I and *Hpa*II; lanes 6 and 7, hybrid DNA (*P. godmani*  $\times$  *P. mareeba*) digested with *Msp*I and *Hpa*II.

quate to demonstrate that the genome of the interspecific hybrid had a much lower level of overall methylation.

To test whether the failure of methylation was specific to this hybrid or was generally associated with hybridization, it was necessary to assay for methylation deficiencies in other hybrids, especially those in natural hybrid zones. DNA from two rock wallaby species (*Petrogale godmani* and *P. mareeba*) and the fertile female hybrid between them (captured in a hybrid zone in north Queensland) was digested with *Msp*I and *Hpa*II (Fig. 6). This hybrid genome showed a clear reduction in methylation, as exemplified by the *Hpa*II digest. Undermethylation of the genome of another rock wallaby hybrid from a different species cross (*P. assimilis*  $\times$  *P. inornata* sterile male hybrid) was also demonstrated by *in situ* nick-translation (see Fig. 2 of Supplementary Information).

There is also evidence that *de novo* chromosome changes, similar to those described here, are common to other interspecies hybrids. Chromosome painting of the *P. assimilis*  $\times$  *P. inornata* hybrid also showed six interchromosomal rearrangements, all involving centromeres (R.J.W.O'N. *et al.*, manuscript in preparation). Our observations, together with previously described hybrid-specific centromeric expansions<sup>10</sup>, indicate that retroelement invasion and amplification in centromeric heterochromatin may be a widespread phenomenon in mammalian hybrids.

It is surprising that levels of methylation were so low in hybrids that are viable and phenotypically normal. Apparently, deficient methylation of the genome and activation of retroelements leading to gross changes in genome structure did not affect the action of developmentally important genes in these interspecific marsupial hybrids. Our results, in conjunction with other observations<sup>18–20</sup>, suggest that there may be alternative mechanisms of genomic methylation. □

#### Methods

**Isolation and characterization of the amplified retroviral element.** *Msp*I and *Hpa*II digests of 10  $\mu$ g of high-molecular-mass DNA from the hybrid male was run on a 3.0% NuSieve (SeaKem) gel in 1  $\times$  TBE (modified) for three days at 4°C. Appropriate bands were then excised from the gel and digested with Agarase (Boehringer). The fragments were cloned into the *Cl*aI site of Bluescript and transformations were subjected to blue–white selection. All sequencing was carried out using a Perkin Elmer Amplicycle sequencing kit and alignments were made using the program Clustal V.

**Southern analysis.** *MspI*, *HpaII*, *BamHI*, and *PstI* digests of 10 µg of high-molecular-mass DNA isolated from blood from tammar wallabies (designation numbers were T151, T126 for Garden Island animals, and 3206, 3275, 3289, 3295, 3221 for Kangaroo Island animals), swamp wallabies (Wb1, Wb2, Wb3, Wb5, Wb6, 110) and the hybrid (BE-1) were electrophoresed on a 0.8% agarose gel in 1 × TBE and then blotted onto Hybond-N+ (Amersham) according to ref. 21. Probe DNA was generated by PCR amplification of plasmid DNA using 50 ng of DNA and 100 ng each of T7 and SP6 primers in a standard 25-µl reaction with 2.5 mM dNTPs, 10 × buffer with 1.5 mM MgCl<sub>2</sub> and *Taq* polymerase. PCR was done (at 94 °C for 30 s, 65 °C for 30 s, 72 °C for 1 min) for 30 cycles. Probe DNA was electrophoresed and digested with Agarase. 50 ng probe DNA was labelled with <sup>32</sup>P-dCTP with a Mega Prime DNA-labelling kit (Amersham) according to the manufacturer's instructions. Blots were hybridized at 65 °C overnight in Church and Gilbert buffer (0.5 M NaHPO<sub>4</sub>, pH 7.0, 1 mM EDTA, pH 8.0, 7% SDS) and washed in 0.1 × SSC, 0.1% at 65 °C. 10 µg genomic DNA isolated from blood from *Petrogale godmani*, *P. mareeba* and *P. godmani* × *P. mareeba* female hybrid was digested with *MspI* and *HpaII*. Digests were run on a 0.8% agarose gel and blotted. Each digest was confirmed to be complete by Southern analysis of cytochrome *b* mitochondrial DNA from *M. eugenii*.

**Fluorescence in situ hybridization.** Chromosomes were prepared from whole blood using standard methods<sup>22</sup> with minor modifications<sup>23</sup>. Plasmid DNA was prepared by the miniprep method and labelled by direct-incorporation PCR. In a 50-µl reaction, 100 ng template was mixed with 100 ng each of primers T7 and SP6, 10 × buffer with 1.5 mM MgCl<sub>2</sub>, 4 µl of 2.5 mM dGAC, 2.4 µl dTTP and 4 µl biotin-16-dUTP (50 nmol µl<sup>-1</sup>) and *Taq* polymerase and subjected to PCR (94 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min) for 40 cycles. The PCR product was purified for further use on a Micron 100 column (Amicon). Chromosome *in situ* suppression hybridization of the probe was done according to standard methods<sup>24</sup> with several modifications. Low-stringency hybridizations were carried out in 8 × SSC, 40% dextran sulphate at 36 °C for 72 h and washed at 45 °C in 0.5 × SSC. High-stringency hybridizations were carried out in 4 × SSC, 40% dextran sulphate at 36 °C for 72 h and washed at 60 °C in 0.1 × SSC. Images were collected on a Leitz microscope with a liquid-cooled Photometrics charge-coupled device camera. Images were processed with NIH Image and registration.

**In situ nick-translation.** Slides were removed from storage at -70 °C and placed in 100% ethanol for 5 min then air-dried. The slide was rinsed in 10 mM Tris-HCl, pH 7.4 and air-dried. 10 units of *HpaII* or *MspI* was mixed into 8 µl 10 × enzyme buffer, placed on the slide and covered with a small coverslip. The reaction was allowed to proceed for 30 min in a moist chamber at 37 °C. Upon completion, the coverslip was removed and the slide placed in 10 mM Tris-HCl, pH 7.4, for 5 min, followed by a standard ethanol series and air-dried. A mix containing 1 µl 10 × DIG labelling MIX (Boehringer), 1 µl nick-translation buffer (BRL), 2 µl DNA polymerase, 0.1 µl β-mercaptoethanol and 5 µl H<sub>2</sub>O was placed on the slide, which was then covered with a coverslip and left at room temperature for 15 min. A time series was conducted on human material for both the digestion and labelling reaction times, with intervals of 5 min from 5 min to 1 h, and optimal times of 30 min and 15 min, respectively, were used for further analysis. The reaction was terminated by rinsing the slide in 10 mM Tris-HCl, pH 7.4, for 5 min. Digoxigenin detection was carried out using anti-digoxigenin-rhodamine fragments according to standard methods<sup>24</sup>.

Received 6 October 1997; accepted 26 February 1998.

- Grant, V. The regulation of recombination in plants. *Cold Spring Harb. Symp. Quant. Biol.* **23**, 337–363 (1958).
- Templeton, A. R. Mechanisms of speciation—a population genetic approach. *Annu. Rev. Ecol. Syst.* **12**, 23–48 (1981).
- Rieseberg, L. H., Van Fossen, C. & Desrochers, A. M. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* **375**, 313–316 (1995).
- Engels, W. R. & Preston, C. R. Formation of chromosome rearrangements by P factors in *D. melanogaster*. *Genetics* **107**, 657–678 (1984).
- Petrov, D. A., Schutzman, J. L., Hartl, D. L. & Lozovskaya, E. R. Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl Acad. Sci. USA* **92**, 8050–8054 (1995).
- Bestor, T. H. & Tycko, B. Creation of genomic methylation patterns. *Nature Genet.* **12**, 363–367 (1996).
- Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
- Arnault, C. & Dufournel, I. Genome and stresses: reactions against aggressions, behavior of transposable elements. *Genetica* **93**, 149–160 (1994).
- Wichman, H. A., Van Den Bussche, R. A., Hamilton, M. J. & Baker, R. J. Transposable elements and the evolution of genome organization in mammals. *Genetica* **86**, 287–293 (1992).

- Smith, M. J., Hayman, D. L. & Hope, R. M. Observations on the chromosomes and reproductive systems of four Macropodine interspecific hybrids (Marsupialia: Macropodidae). *Aust. J. Zool.* **27**, 959–972 (1979).
- Toder, R. *et al.* Comparative chromosome painting between two marsupials: origins of an XX/XY<sub>2</sub> sex chromosome system. *Mamm. Genome* **8**, 418–422 (1997).
- Holland, J. *et al.* Rapid evolution of RNA genomes. *Science* **215**, 1577–1585 (1982).
- Graur, D. Pattern of nucleotide substitution and the extent of purifying selection in retroviruses. *J. Mol. Evol.* **21**, 221–231 (1985).
- Jaenisch, R., Schnieke, A. & Herbers, K. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. *Proc. Natl Acad. Sci. USA* **82**, 1451–1455 (1985).
- Sunkel, C. E. & Coelho, P. A. The elusive centromere: sequence divergence and functional conservation. *Curr. Opin. Genet. Dev.* **5**, 756–767 (1995).
- Curcio, M. J. & Morse, R. H. Tying together integration and chromatin. *Trends Genet.* **12**, 436–438 (1996).
- Loebel, D. A. & Johnston, P. G. Analysis of DNase I activity and methylation of active and inactive X chromosomes of kangaroos (*Macropus robustus*) by *in situ* nick translation. *Chromosoma* **102**, 81–87 (1993).
- Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
- Jacobsen, S. E. & Meyerowitz, E. M. Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* **277**, 1100–1103 (1997).
- Weiss, A., Keshet, I., Razin, A. & Cedar, H. DNA demethylation *in vitro*: involvement of RNA. *Cell* **86**, 709–718 (1996).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
- Schemp, P. W. & Meer, B. Cytologic evidence for three human X-chromosomal segments escaping inactivation. *Hum. Genet.* **63**, 171–174 (1983).
- Toder, R., Wilcox, S. A., Smithwick, M. & Graves, J. A. M. The human/mouse imprinted genes *IGF2*, *H19*, *SNRPN*, and *ZNF127* map to two conserved autosomal clusters in a marsupial. *Chromosome Res.* **4**, 295–300 (1996).
- Lichter, P., Cremer, T., Borden, J., Manuelidis, L. & Ward, D. C. Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridization using recombinant DNA libraries. *Hum. Genet.* **80**, 224–234 (1988).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

**Acknowledgements.** We thank M. Eldridge for swamp wallaby, tammar wallaby, *Petrogale* sp. and hybrid material, M. Renfree for tammar wallaby material, and I. Barbieri for cell preparations from BE-1.

Correspondence and requests for materials should be addressed to R.J.W.O.N. (e-mail: [racho@princeton.edu](mailto:racho@princeton.edu)). Genbank accession number of KERV-1 sequence is AF044909.

## Hypothalamic CART is a new anorectic peptide regulated by leptin

Peter Kristensen, Martin E. Judge, Lars Thim, Ulla Ribel, Kennet N. Christjansen, Birgitte S. Wulff, Jes T. Clausen, Per B. Jensen\*, Ole D. Madsen\*, Niels Vrang†, Philip J. Larsen† & Sven Hastrup

Health Care Discovery, Novo Nordisk A/S, Novo Alle, 6A1.47, DK-2880 Bagsvaerd, Denmark

\* Hagedorn Research Institute, Niels Steensevej 6, DK-2820 Gentofte, Denmark

† Department of Anatomy, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N., Denmark

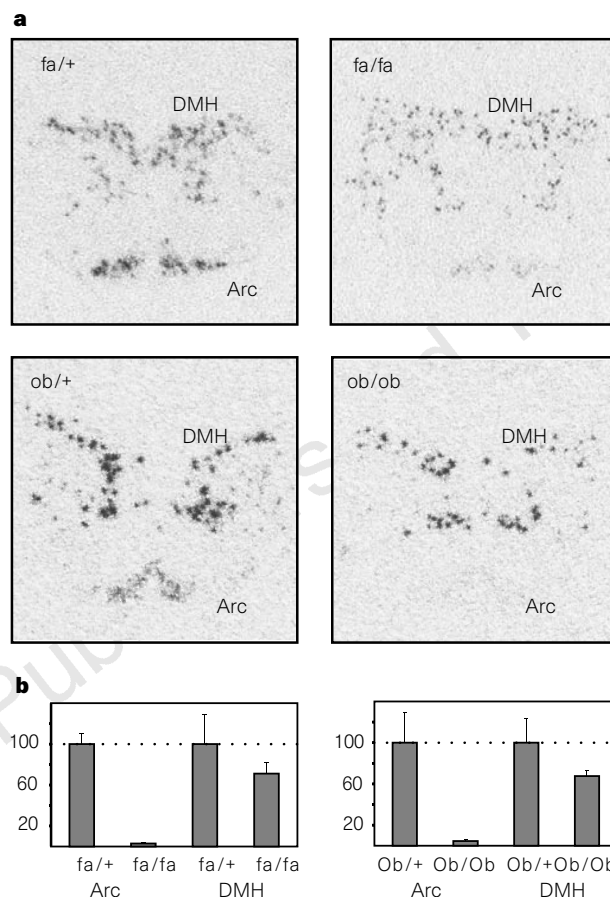
The mammalian hypothalamus strongly influences ingestive behaviour through several different signalling molecules and receptor systems<sup>1–4</sup>. Here we show that CART (cocaine- and amphetamine-regulated transcript), a brain-located peptide<sup>5–8</sup>, is a satiety factor and is closely associated with the actions of two important regulators of food intake, leptin and neuropeptide Y. Food-deprived animals show a pronounced decrease in expression of CART messenger RNA in the arcuate nucleus. In animal models of obesity with disrupted leptin signalling, CART mRNA is almost absent from the arcuate nucleus. Peripheral administration of leptin to obese mice stimulates CART mRNA expression. When injected intracerebroventricularly into rats, recombinant CART peptide inhibits both normal and starvation-induced feeding, and completely blocks the feeding response induced by neuropeptide Y. An antiserum against CART increases feeding in normal rats, indicating that CART may be an endogenous

# inhibitor of food intake in normal animals.

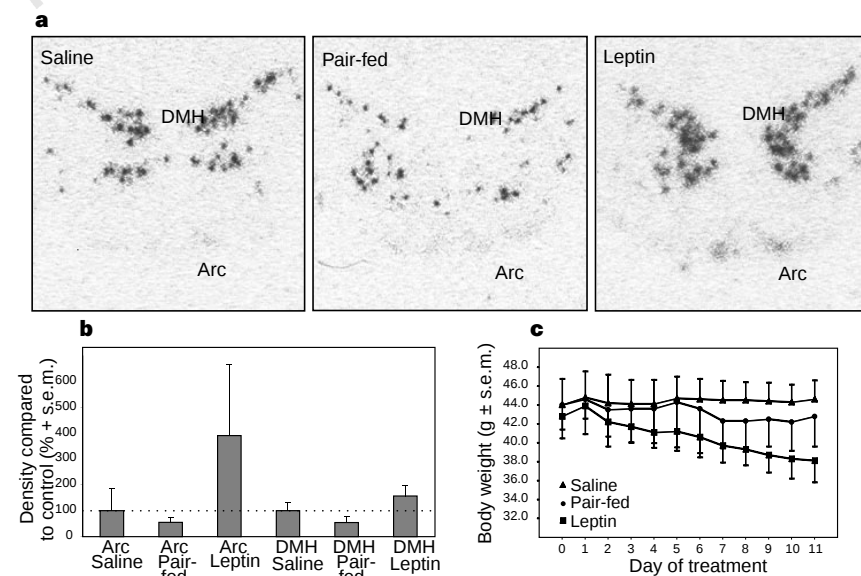
Using *in situ* hybridization with <sup>35</sup>S-labelled antisense RNA probes, we analysed the expression of CART mRNA in the hypothalamus of genetically obese rodents. Compared with lean controls, obese Zucker rats (*fa/fa* rats) showed a significant decrease in CART mRNA levels in the arcuate nucleus. Also, obese Zucker rats exhibited a possible decrease in CART mRNA levels over the dorsomedial and lateral hypothalamic areas, although this decrease was not significant (Fig. 1a, b). Obese (*ob/ob*) mice, which lack functional leptin showed a similar downregulation of CART mRNA expression in the arcuate nucleus as compared with heterozygous (*ob/+*) mice. Obese mice had very low levels of expression of CART mRNA when no leptin signal was present (Fig. 1c, d). Likewise, *ob/ob* mice also showed a significant but less pronounced downregulation of CART mRNA expression in the dorsomedial and lateral hypothalamic area.

To examine whether hypothalamic CART mRNA expression is regulated by circulating leptin, we injected *ob/ob* mice with recombinant leptin intraperitoneally once daily for 10 days. The group of leptin-treated *ob/ob* mice had significantly lower food intake compared with the group treated with saline ( $20 \pm 3$  g per day versus  $43 \pm 3$  g per day, mean  $\pm$  s.d.). Expression of CART mRNA was restored in the arcuate nucleus and increased in the lateral hypothalamus of the leptin-treated mice (Fig. 2). In contrast, a group of *ob/ob* mice pair-fed with the leptin-treated group showed a small decrease in the CART mRNA signal over both the arcuate nucleus and the dorsomedial hypothalamus (DMH), showing that altered food intake following leptin treatment had no effect on CART mRNA expression *per se*. When normal rats were fasted for 24 or 48 h, levels of CART mRNA in the arcuate nucleus decreased (Fig. 3), and a similar change, which did not quite reach statistical significance ( $P = 0.057$ ), was observed in the DMH.

Synthesis of neuropeptide Y (NPY) in the arcuate nucleus, and secretion of NPY in the paraventricular nucleus, increases after fasting<sup>9,10</sup>. Similarly, levels of arcuate NPY are increased in both the *fa/fa* rat<sup>11</sup> and the *ob/ob* mouse<sup>12</sup>. We therefore investigated the effect of CART on feeding both in normal rats and in rats in which feeding had been stimulated by intracerebroventricular (i.c.v.) administration of NPY. Several post-translationally processed forms of CART are produced when the rat CART complementary DNA is expressed in yeast. One of these forms has an amino terminus identical to a partial peptide sequence identified during the purification of somatostatin from sheep hypothalamus<sup>13</sup>, indicating that this protein is a physiologically relevant hypothalamic neuropeptide



**Figure 1** Expression of CART mRNA in obese rats and mice. **a**, *In situ* hybridization of CART mRNA in sections from the hypothalamus of *Fa/+* and *fa/fa* Zucker rats and *ob/+* and *ob/ob* mice. Dark areas indicate mRNA expression. Note the downregulation of CART mRNA in the arcuate nucleus (Arc) and the dorsomedial hypothalamus (DMH) in both homozygous animal strains. **b**, Quantification by image analysis of areas of CART mRNA expression with a mean density (percentage of control, y-axis) above threshold for two sections from six animals in each group. Films were exposed for 12 days (rat sections) or 24 days (mouse sections). Statistical analysis: *Fa/+* versus *fa/fa*, Arc,  $P = 0.004$ , and DMH,  $P = 0.154$ , *ob/+* versus *ob/ob*, Arc,  $P = 0.002$ , and DMH,  $P = 0.002$ , using a Mann-Whitney U-test.



**Figure 2** CART mRNA expression is induced by leptin. **a**, Treatment of *ob/ob* mice with daily intraperitoneal injections of 150  $\mu$ g recombinant leptin (for 10 days) leads to reappearance of CART mRNA expression in the arcuate nucleus (Arc) and to an increase of CART mRNA expression in the dorsomedial hypothalamus (DMH) of the leptin-treated mice. **b**, Quantitative image analysis of areas of CART mRNA expression with a mean density above threshold from two sections from six animals in each group. Film was exposed for 21 days. **c**, Changes in body weight during treatment. Expression in the arcuate nucleus: leptin versus saline,  $P = 0.041$ ; leptin versus pair-fed,  $P = 0.004$ . Expression in the dorsomedial hypothalamus: leptin versus saline,  $P = 0.053$ ; leptin versus pair-fed,  $P = 0.002$ , using a Mann-Whitney U-test.

(named CART(55–102), the numbers derive from the predicted signal-peptide-cleavage site in the long form of CART<sup>5</sup>. Central administration of recombinant CART(55–102) inhibited feeding both in non-fasted rats (Fig. 4a) and in NPY-challenged rats (Fig. 4b). Centrally administered NPY and CART peptides have different pharmacokinetic profiles, with NPY having the most sustained effect. Thus, in both normal and NPY-stimulated rats, the inhibitory effect of CART was strongest during the first hour after injection, and the stimulatory effect of injected NPY was suppressed by CART in a dose-dependent manner.

CART was centrally administered to rats that had been fasted for 24 h. Food intake during the first hour after administration of CART was measured. 1 µg CART inhibited feeding by 59% and 2 µg CART inhibited feeding by 68% (food intake (mean ± s.e.m.): control, 11.2 ± 1.7 g; 1 µg CART, 4.6 ± 1.6 g; 2 µg CART, 3.6 ± 1.7 g;  $P = 0.014$ , ANOVA). In fasted mice, i.c.v. injection of 2 µg CART(61–102) or CART(55–102) completely inhibited liquid-diet intake in a brief feeding test. This effect was lost when we disrupted the secondary structure of CART by reduction and pyridylethylation, which led to breakage of the three disulphide bonds in the molecule (L.T. *et al.*, unpublished observations).

In addition to having inhibitory effects on feeding, CART treatment induced some movement-associated tremor, but the rats were clearly capable of eating (Fig. 4b, 1 µg CART + 5 µg NPY group). I.c.v. injections of 1 or 2 µg CART did not significantly alter spontaneous locomotor activity when individual rats were tested for 1 h in isolated activity-test chambers (Photobeam interruptions

(mean ± s.e.m.): control, 3,498 ± 289; 1 µg CART, 4,097 ± 387; 2 µg CART, 3,458 ± 442;  $P = 0.425$ , analysis of variance (ANOVA).

To determine whether an endogenous tone of CART signalling exists in the rat hypothalamus, we investigated whether an anti-serum against CART affects normal night-time feeding behaviour. This was done in a crossover design, in which all animals on separate days (in mixed order) received a central injection of 5 µl anti-CART antiserum or a similar amount of preimmune serum from the same rabbit. Compared with preimmune serum, anti-CART antiserum significantly increased night-time feeding (Table 1). The effect of treatment with anti-CART antiserum was most pronounced during the last part of the feeding session, indicating a blockade of an endogenously secreted satiety factor. Furthermore, when we precipitated the immunoglobulins that specifically recognize CART from the anti-CART antiserum by incubation with CART-coupled Sepharose, this CART-cleared antiserum had no effect on feeding, as compared with the effects of similarly treated preimmune serum (Table 1). Finally, when we preincubated the CART peptide with the antibody, the inhibitory effect of CART on food intake was abolished (results not shown).

To study the role of the CART peptide in the hypothalamus further and to characterize the CART antiserum, we localized the CART peptide by immunohistochemistry. There was prominent staining in neurons of the ventrolateral part of the arcuate nucleus (Fig. 5a) and strong staining in the external zone of the median eminence (Fig. 5c). The pattern of staining is different from that found when staining for NPY. As the pattern of CART mRNA

**Table 1** CART antiserum increases night-time food intake

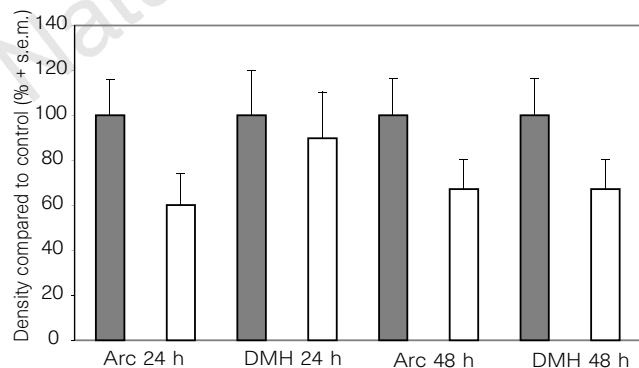
| Serum                  | Food intake<br>0–2 h | Food intake<br>2–12 h | Food intake<br>0–12 h |
|------------------------|----------------------|-----------------------|-----------------------|
| Preimmune              | 3.2 ± 0.5            | 7.9 ± 1.1             | 11.1 ± 1.3            |
| Anti-CART              | 2.6 ± 0.3            | 12.1 ± 0.7†‡          | 14.7 ± 0.8†           |
| Preimmune<br>Absorbed* | 3.0 ± 0.6            | 9.7 ± 1.2             | 12.7 ± 1.0            |
| Anti-CART<br>Absorbed* | 3.1 ± 0.6            | 8.8 ± 1.1             | 11.9 ± 0.8            |

Animals received an i.c.v. injection of preimmune or anti-CART antiserum (see Methods).

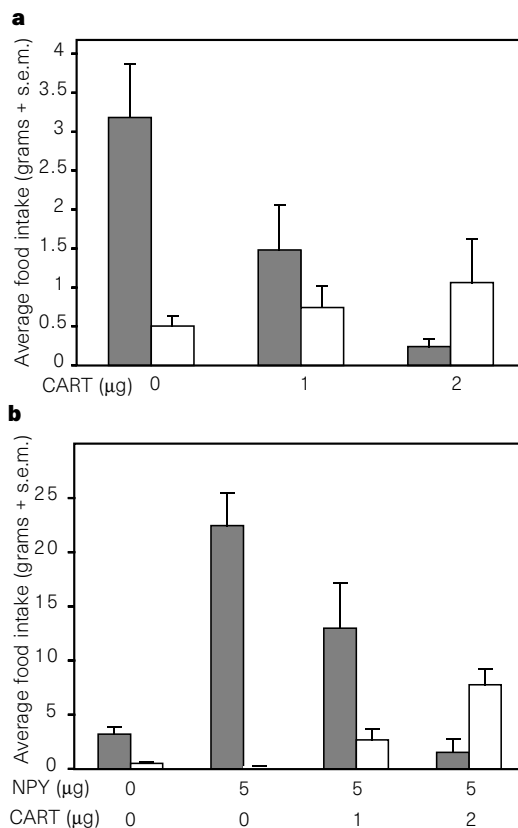
\* Batch absorption with Sepharose-coupled CART.

†  $P < 0.05$  versus preimmune serum.

‡  $P < 0.05$  versus anti-CART serum absorbed (ANOVA followed by Fisher probability of least squares deviation. Food intake is measured in g ± s.e.m. ( $n = 10–15$ )).



**Figure 3** Quantification of CART mRNA *in situ* hybridization in sections from the hypothalamus of non-fasted (filled bars) and fasted (open bars) Wistar rats. The figure shows quantitative image analysis of the area of CART mRNA expression with a mean density above threshold in the arcuate nucleus (Arc) and the dorsomedial hypothalamus (DMH) (24-h fasting: two sections per animal, 6 animals; 48-h fasting: three sections per animal, six animals). In the same experiment, NPY mRNA levels were increased by sevenfold in the arcuate nucleus (result not shown). Control versus fasted: arcuate nucleus, 24-h fast,  $P = 0.03$ , and 48-h fast,  $P = 0.026$ ; DMH, 24-h,  $P = 0.51$ , and 48-h fast,  $P = 0.057$ , using a Mann-Whitney U-test.



**Figure 4** CART reduces food intake. **a**, Effect of recombinant CART peptide, injected intracerebroventricularly into non-fasted rats, on feeding during the first (filled bars) and second (open bars) hours after injection. **b**, Inhibition of feeding in rats receiving different amounts of CART peptide followed by 5 µg porcine NPY. ANOVA of effect of CART: CART alone, first hour,  $P = 0.058$ , second hour,  $P = 0.57$ ; CART + NPY, first hour,  $P = 0.0015$ , second hour,  $P = 0.0005$ .



hybridization<sup>5</sup> resembles that found for the proopiomelanocortin (POMC) precursor peptide mRNA, we stained adjacent sections with antibodies recognizing the amidated hinge peptide of the POMC precursor. The CART-staining pattern found was different from that of POMC (Fig. 5b). Interestingly, CART immunostaining could also be found in terminals in the periventricular part of the paraventricular nucleus (Fig. 5d), indicating that CART may have a neuromodulatory effect in this nucleus, which is the site of action for NPY-stimulated feeding<sup>9</sup>.

We have shown here that the expression of CART mRNA in the arcuate nucleus depends on leptin signalling, as disruption of either circulating leptin levels (in the *ob/ob* mouse) or the leptin receptor (in the *fa/fa* rat) leads to decreased expression of CART. A reverse relationship exists for NPY. Thus, although there were no changes in food ingestion and body weight in the NPY-knockout mouse<sup>14</sup>, crossbreeding with leptin-deficient animals reduced the obesity of the *ob/ob* animals<sup>15</sup>. These and other findings<sup>16</sup> indicate that circulating leptin is normally a repressor of NPY production in the arcuate nucleus. NPY and CART may thus be two of the factors that balance the effects of leptin on feeding through the hypothalamus. The finding of a hypothalamic leptin-dependent satiety factor may also explain the initial elevation of sensitivity to leptin in NPY-knockout mice<sup>14</sup>.

We propose that leptin-mediated suppression of food intake is controlled by a balanced reduction and induction of NPY and

CART, respectively. A putative receptor mediating the neuromodulatory action of CART could thus represent an interesting therapeutic target for the control of obesity. □

## Methods

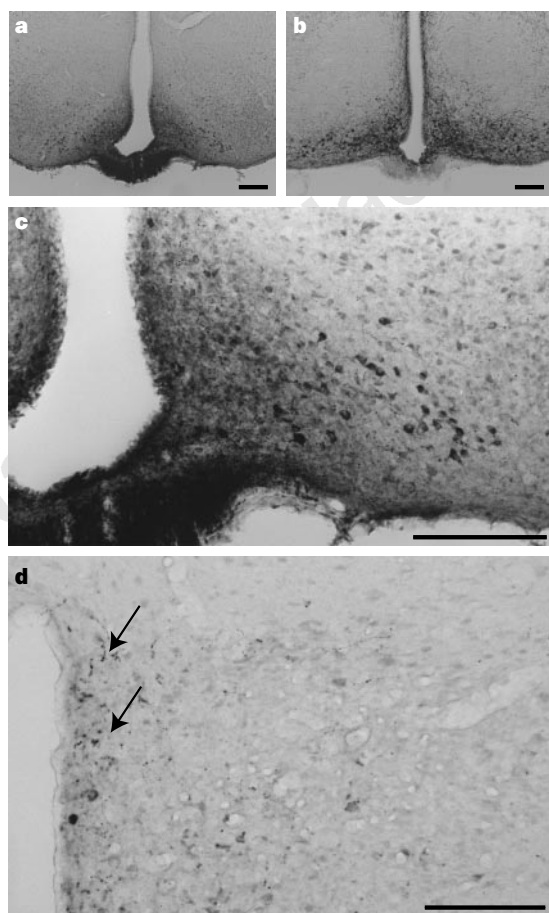
**CART-peptide expression and purification.** The cDNA encoding rat CART(1–102) was inserted in the expression plasmid pAK405. We used a yeast expression system previously optimized for insulin production<sup>17</sup>. Following fermentation, various CART fragments were purified by cationic-exchange chromatography using an SP–Sephacolumn (Pharmacia) followed by preparative HPLC on a C4 column (Vydac). Peptide structure was determined by amino-acid-sequence analysis (Applied Biosystems 494 Protein Sequencer) and electrospray mass-spectrometry analysis (Sciex API III LC/MS/MS system).

**Animal experiments.** For i.c.v. injection of CART, male Wistar rats were implanted with guide cannulae in the lateral ventricle and allowed to recover for 4–8 days, before screening for functional cannulae by injection of 5 µg porcine NPY (pNPY). Non-responders were discarded, and the remaining rats sorted into response-equivalent groups of five rats. The rats had access to the test food, a mash made from a 2:1 mixture of water and dry chow, for 24 h before receiving an i.c.v. injection of 5 µl phosphate-buffered saline (PBS) or 1 or 2 µg CART(55–102) in PBS. Food consumption was measured for 2 h. To examine the effect of CART on NPY-induced feeding, we followed the first i.c.v. injection immediately with 5 µl of saline or 5 µg of pNPY. Locomotor activity was measured for 1 h in individual non-fasted rats after i.c.v. injection of PBS or CART (1 or 2 µg) by recording infrared photobeam interruptions in isolated chambers. Statistical analysis was done using ANOVA.

To analyse the effect of CART antiserum on night-time feeding, we equipped adult male Wistar rats with guide cannulae in the lateral ventricle. These rats were housed individually in metabolic cages, under a light cycle of 12 lights on:12 lights off, and were allowed to acclimatize for 10 days. On separate days, animals were injected intracerebroventricularly with 5 µl of either anti-CART antiserum (AB 2055A) or preimmune serum (from the same rabbit) immediately before the lights were turned off, thus serving as their own controls. A period of 2 days was allowed between injections. Food intake was monitored after 2 h and the following morning (after 12 h). For control experiments, both antisera were incubated with CNBr–Sephacolumn-coupled CART overnight at 4 °C; Sepharose particles were then removed by centrifugation. For leptin treatment, *ob/ob* mice were injected once daily intraperitoneally with 150 µg purified human leptin. When analysed using ANOVA (Student–Newman–Keuls method), the CART expression levels found in both the arcuate nucleus and the DMH of the leptin-treated group were different from levels in both the pair-fed and the saline-treated groups ( $P < 0.05$ ).

**In situ hybridization.** *In situ* hybridization analysis was performed on cryostat sections<sup>18</sup> using antisense RNA probes against the rat CART cDNA (base pairs 226–411; Genbank accession number U10071). Post-hybridization washes were performed at 62 °C and 67 °C in 50% formamide. After hybridization, sections were exposed on Beta-Max film (Amersham). Images were scanned using a 2,000 d.p.i. slide scanner and analysed using the NIH–Image software by an observer blinded to the treatment/genetic background of individual animals. There was no signal when the corresponding sense RNA probe was used as a control. Additional hybridization with antisense RNA probes corresponding to base pairs 17–225 of the cDNA showed an identical pattern of hybridization to that observed with base pairs 226–411.

**Immunohistochemistry.** The cDNA encoding CART(1–102) D(27–39) (ref. 5) was inserted into the glutathione-S-transferase gene fusion system plasmid pGEX-2T (Pharmacia) and the fusion protein was expressed in *Escherichia coli*. The protein was purified on glutathione–Sephacolumn and used for immunizing NZW rabbits for the production of antiserum (2055A). Antiserum recognizing the amidated hinge peptide of the POMC precursor (HPN-6970) was as described<sup>19</sup>. Free-floating sections from brains of perfusion-fixed adult male Wistar rats were incubated for 24 h in primary antibodies 2055A (diluted 1:1,000) or HPN-6970 (diluted 1:2,000), and bound antibodies were detected using the tyramide amplification of avidin/biotin/peroxidase method<sup>20</sup>. Control sections for single-antigen immunohistochemistry were routinely processed by either omitting or replacing the primary antiserum with



**Figure 5** CART is present in neurons of the hypothalamic arcuate nucleus. **a**, Immunohistochemical localization of CART peptide in the rat hypothalamus using rabbit antibodies. **b**, Comparison of CART localization with localization of the POMC precursor peptide. **c**, Neurons of the lateral arcuate nucleus show strong staining, as does the terminal field in the external zone of the median eminence. **d**, Nerve terminals of the paraventricular nucleus show CART immunostaining. Scale bars, 100 µm.

an equivalent concentration of rabbit preimmune serum. We found identical staining patterns when using an antibody raised against the insert peptide of the long form<sup>5</sup> of rat CART.

Received 10 December 1997; accepted 17 March 1998.

- Clark, J. T., Kalra, P. S., Crowley, W. R. & Kalra, S. P. Neuropeptide Y and human pancreatic polypeptide stimulate feeding behavior in rats. *Endocrinology* **115**, 427–429 (1984).
- Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263–1271 (1995).
- Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165–168 (1997).
- Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–141 (1997).
- Douglass, J., McKinzie, A. A. & Couceyro, P. PCR differential display identifies a rat brain mRNA that is transcriptionally regulated by cocaine and amphetamine. *J. Neurosci.* **15**, 2471–2481 (1995).
- Gautvik, K. M. *et al.* Overview of the most prevalent hypothalamus-specific mRNAs, as identified by directional tag PCR subtraction. *Proc. Natl Acad. Sci. USA* **93**, 8733–8738 (1997).
- Couceyro, P. R., Koylu, E. O. & Kuhar, M. J. Further studies on the anatomical distribution of CART by *in situ* hybridization. *J. Chem. Neuroanat.* **12**, 229–241 (1997).
- Douglass, J. & Daoud, S. Characterization of the human cDNA and genomic DNA encoding CART: a cocaine- and amphetamine-regulated transcript. *Gene* **169**, 241–245 (1996).
- Kalra, S. P., Dube, M. G., Sahu, A., Phelps, C. P. & Kalra, P. S. Neuropeptide Y secretion increases in the paraventricular nucleus in association with increased appetite for food. *Proc. Natl Acad. Sci. USA* **88**, 10931–10935 (1991).
- White, J. D. & Kershaw, M. Increased hypothalamic Neuropeptide Y expression following food deprivation. *Mol. Cell. Neurosci.* **1**, 41–48 (1990).
- Sanacora, G., Kershaw, M., Finkelstein, J. A. & White, J. D. Increased hypothalamic content of preneuropeptide Y messenger ribonucleic acid in genetically obese Zucker rats and its regulation by food deprivation. *Endocrinology* **127**, 730–737 (1990).
- Wilding, J. P. *et al.* Increased neuropeptide-Y messenger ribonucleic acid (mRNA) and decreased neuropeptide Y mRNA in the hypothalamus of the obese (ob/ob) mouse. *Endocrinology* **132**, 1939–1944 (1993).
- Spies, J., Villarreal, J. & Vale, W. Isolation and sequence analysis of a somatostatin-like polypeptide from ovine hypothalamus. *Biochemistry* **20**, 1982–1988 (1981).
- Erickson, J. C., Clegg, K. E. & Palmiter, R. D. Sensitivity to leptin and susceptibility to seizures of mice lacking neuropeptide Y. *Nature* **381**, 415–421 (1996).
- Erickson, J. C., Holloper, G. & Palmiter, R. D. Attenuation of the obesity syndrome of ob/ob mice by the loss of neuropeptide Y. *Science* **274**, 1704–1707 (1996).
- Wang, Q. *et al.* Interactions between leptin and hypothalamic neuropeptide Y neurons in the control of food intake and energy homeostasis in the rat. *Diabetes* **46**, 335–341 (1997).
- Kjeldsen, T. *et al.* A removable spacer peptide in an alpha-factor-leader/insulin precursor fusion protein improves processing and concomitant yield of the insulin precursor in *Saccharomyces cerevisiae*. *Gene* **170**, 107–112 (1996).
- Kristensen, P., Eriksen, J. & Dano, K. Localization of urokinase-type plasminogen activator messenger RNA in the normal mouse by *in situ* hybridization. *J. Histochem. Cytochem.* **39**, 341–349 (1991).
- Fenger, M. Proopiomelanocortin. *Scand. J. Clin. Lab. Invest.* **50**, 229–245 (1990).
- Berghorn, K. A., Bonnett, J. H. & Hoffman, G. E. cFos immunoreactivity is enhanced with biotin amplification. *J. Histochem. Cytochem.* **42**, 1635–1642 (1994).

**Acknowledgements.** We thank J. Kuijper for the gift of recombinant leptin; I. Diers, A. S. Andersen and P. E. Nielsen for production of CART in yeast and characterization of the purified peptide; and E. Bentsen, B. Jørgensen, W. Listov-Saabye, L.-L. Kruse, J. Mandelbaum, H. Petersen, L. Priskorn and S. Kryger for technical assistance. P.B.J. and O.D.M. were supported by the Danish Cancer Society and the Danish National Research Foundation. Hagedorn Research Institute is a basic research component of Novo Nordisk A/S. P.J.L. and N.V. were supported by the Danish Medical Research Council, the Danish Diabetes Association and the Novo Nordisk Foundation.

Correspondence and requests for materials should be addressed to P.K. (e-mail: pekr@novo.dk).

## Dramatic decreases in brain reward function during nicotine withdrawal

Mark P. Epping-Jordan\*, Shelly S. Watkins, George F. Koob & Athina Markou

The Scripps Research Institute, Department of Neuropsychopharmacology, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

Tobacco smoking is a worldwide public health problem. In the United States alone, over 400,000 deaths and \$50 billion in medical costs annually are directly attributed to smoking<sup>1</sup>. Accumulated evidence indicates that nicotine is the component of tobacco smoke that leads to addiction<sup>2</sup>, but the means by which nicotine produces addiction remain unclear. Nicotine is less effective as a positive reinforcer than other drugs of abuse in non-dependent animals<sup>3</sup>. Nevertheless, nicotine-withdrawal

symptoms, including depressed mood, anxiety, irritability and craving<sup>4,5</sup> in dependent subjects may contribute to the addictive liability of nicotine<sup>6,7</sup>. We show here that spontaneous nicotine withdrawal in rats resulted in a significant decrease in brain reward function, as measured by elevations in brain reward thresholds, which persisted for four days. Further, systemic injections of a competitive nicotinic-receptor antagonist<sup>8</sup> led to a dose-dependent increase in brain reward thresholds in chronic nicotine-treated rats. The decreased function in brain reward systems during nicotine withdrawal is comparable in magnitude and duration to that of other major drugs of abuse<sup>9–13</sup>, and may constitute an important motivational factor that contributes to craving, relapse and continued tobacco consumption in humans<sup>7</sup>.

Abstinence syndromes typically incorporating somatic signs have long been observed in animals after termination of chronic administration of opiates, sedative-hypnotics and ethanol<sup>14–16</sup>; however, only recently, a similar abstinence syndrome has been seen after termination of chronic nicotine administration in the rat<sup>17,18</sup>. The rodent somatic nicotine abstinence syndrome occurs spontaneously after termination of chronic nicotine administration, and can be precipitated by nicotinic<sup>18,19</sup> or opiate<sup>20</sup> receptor antagonists<sup>7</sup>; it is reversible by subsequent nicotine administration<sup>17,18</sup>. However, it is unlikely that somatic symptoms are as critical to the motivational aspects of dependence as are changes in affective state and anionality.

Here we investigate whether the nicotine abstinence syndrome, similar to withdrawal from other addictive drugs, is also characterized by decreases in brain reward function, as measured by elevations in intracranial self-stimulation (ICSS) brain reward thresholds<sup>10,11</sup>. We examined abstinence effects after chronic subcutaneous nicotine administration delivered by osmotic minipumps dispensing 9.0 mg kg<sup>-1</sup> per day of nicotine hydrogen tartrate salt (3.16 mg kg<sup>-1</sup> d<sup>-1</sup> nicotine base). This dose maintains stable plasma nicotine levels (~44 ng ml<sup>-1</sup>; ref. 21) comparable to those reported for smokers consuming 30 cigarettes daily (40–42 ng ml<sup>-1</sup>)<sup>22</sup>. Rats with chronic bipolar stimulating electrodes, demonstrating stable baseline ICSS reward thresholds, were prepared with osmotic minipumps containing either nicotine or saline. All rats were tested in a discrete-trial current-intensity ICSS reward threshold paradigm<sup>10,11</sup> once daily during the chronic nicotine-administration period. For the spontaneous nicotine withdrawal experiment, ICSS reward thresholds and somatic withdrawal signs were assessed after surgical removal of the minipumps; for the precipitated nicotine withdrawal experiment, these same measures were assessed after subcutaneous injections of saline or various doses of the competitive nicotinic receptor antagonist dihydro-β-erythroidine (DHβE).

During chronic nicotine administration, there were no significant changes in baseline reward thresholds for any of the groups (data not shown). The lack of effect of acute nicotine on ICSS reward thresholds is not surprising because minipump infusion rates for nicotine were relatively slow compared to injection rates shown previously to lower ICSS reward thresholds<sup>23,24</sup> or increase dopamine release or energy metabolism in the nucleus accumbens<sup>25,26</sup>. There were also no significant differences between somatic withdrawal signs observed one day before minipump implantation and those observed on the sixth or seventh day of nicotine administration (data not shown).

After removal of the minipumps in the spontaneous withdrawal groups, significant elevations in ICSS reward thresholds, which peaked at 6–8 h and exceeded 140% of baseline values, were observed in nicotine-treated rats (Fig. 1a). Statistically significant differences in reward thresholds between saline- and nicotine-treated rats emerged 4 h after minipump removal and persisted for 4 days. Further, reward threshold elevations in nicotine-treated rats were highly consistent both within and between animals across repeated tests. Thresholds of saline-treated rats did not differ significantly

\* Present address: Glaxo Wellcome Experimental Research S.A., Institut de Biologie Cellulaire et de Morphologie (IBCM), Université de Lausanne, Rue du Bugnon 9, CH-1005 Lausanne, Switzerland.

from baseline values at any time point. Mean thresholds for nicotine-treated rats returned to baseline levels by the fifth day (Fig. 1a), and thresholds for all individual nicotine-treated rats returned to baseline levels within 16 days of minipump removal. These results indicate that abstinence after chronic administration of nicotine induces significant elevations in ICSS reward thresholds, similar to those observed during withdrawal from other addictive drugs.

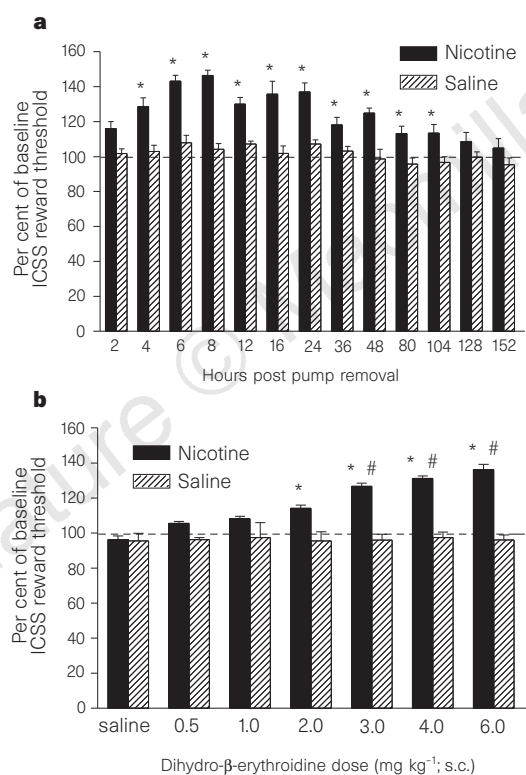
Subcutaneous injection of DH $\beta$ E resulted in statistically significant dose-dependent elevations in ICSS reward thresholds in rats receiving chronic nicotine, but not in saline-treated subjects (Fig. 1b). ICSS reward thresholds on days between DH $\beta$ E drug-test days did not differ from baseline thresholds in either nicotine- or saline-treated rats. These results indicate that nicotine withdrawal, as measured by ICSS reward threshold elevations, can be precipitated by DH $\beta$ E, and suggest that threshold elevations during nicotine withdrawal are mediated, at least in part, by changes in nicotinic acetylcholine receptor function and/or expression.

After removal of the minipumps, statistically significant increases in somatic withdrawal signs were evident in rats undergoing spontaneous nicotine withdrawal. The overall number of somatic withdrawal signs in nicotine-treated rats remained significantly elevated above the overall number for saline-treated rats for 48 h, whereas saline-treated rats showed no somatic signs of withdrawal at any time point (Fig. 2a). The most prominent feature of the

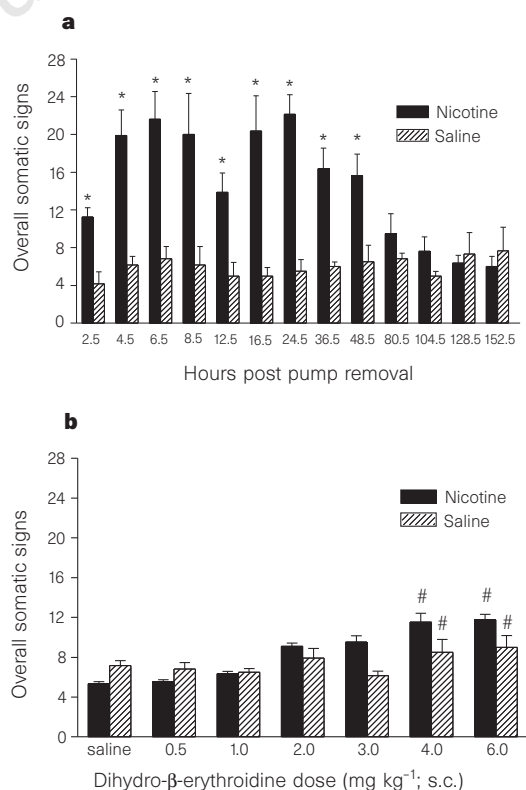
somatic nicotine withdrawal syndrome described here was a statistically significant increase in abdominal constrictions (described as gasps and/or writhes<sup>17,18</sup>) and eye blinks (Table 1), similar to previous reports of spontaneous nicotine withdrawal<sup>17</sup> with a time course longer than in the original report<sup>17</sup> but more consistent with the time course of another report<sup>18</sup>. Hence, our results confirm the existence of a somatic nicotine withdrawal syndrome, which occurs spontaneously in rats after termination of chronic nicotine administration.

Subcutaneous injection of DH $\beta$ E produced statistically significant increases in somatic withdrawal signs at the two highest doses tested. However, the overall number of somatic signs did not differ between nicotine- and saline-treated rats, and was substantially lower than the number of overall somatic signs observed during spontaneous nicotine withdrawal (Fig. 2b). These results provide evidence of a dissociation between ICSS reward threshold elevations and somatic withdrawal signs precipitated by DH $\beta$ E.

The effect of nicotine withdrawal in elevating ICSS reward thresholds, as described here, is similar to the effects of withdrawal from amphetamine<sup>9</sup>, cocaine<sup>10</sup>, opiates<sup>12</sup> and ethanol<sup>13</sup>. The dissociation of the elevations in ICSS reward threshold from the somatic withdrawal syndrome suggests that elevations in ICSS thresholds may provide a more sensitive index of nicotine withdrawal than overt somatic symptoms. These results, showing



**Figure 1** Intracranial self-stimulation (ICSS) reward thresholds in rats (expressed as a percentage of the mean ( $\pm$ s.e.m.) baseline reward threshold). ICSS reward thresholds during **a**, spontaneous withdrawal after termination of chronic administration of nicotine hydrogen tartrate (9.0 mg kg<sup>-1</sup> d<sup>-1</sup> for 7 d) ( $n$  = 8) or saline ( $n$  = 6); or **b**, withdrawal precipitated by dihydro- $\beta$ -erythroidine (DH $\beta$ E), a competitive nicotinic-receptor antagonist, during chronic administration of nicotine hydrogen tartrate (9.0 mg kg<sup>-1</sup> d<sup>-1</sup>) ( $n$  = 9) or saline ( $n$  = 6). Asterisks indicate statistically significant differences ( $P$  < 0.05) between nicotine- and saline-treated groups: **a**, after removal of the minipumps, or **b**, after DH $\beta$ E injection, compared to the effects of saline (0.0 mg kg<sup>-1</sup> DH $\beta$ E, s.c.). In **b**, hash symbols indicate statistically significant increases ( $P$  < 0.05) in ICSS thresholds compared to those after injection of 2.0 mg kg<sup>-1</sup> DH $\beta$ E in nicotine-treated rats, indicating a dose-dependent effect of DH $\beta$ E. s.c., Subcutaneous injection.



**Figure 2** Overall somatic withdrawal signs. Withdrawal signs during **a**, spontaneous withdrawal after termination of chronic nicotine (9.0 mg kg<sup>-1</sup> d<sup>-1</sup> for 7 days) ( $n$  = 8) or saline ( $n$  = 6) administration; or **b**, withdrawal precipitated by DH $\beta$ E during chronic nicotine (9.0 mg kg<sup>-1</sup> d<sup>-1</sup>) ( $n$  = 9) or saline ( $n$  = 6) administration. In **a**, asterisks indicate statistically significant differences in overall somatic withdrawal signs ( $P$  < 0.05) between saline and nicotine groups after minipump removal; in **b**, hash symbols indicate statistically significant increases ( $P$  < 0.05) in overall somatic withdrawal signs compared to 0.0 mg kg<sup>-1</sup> DH $\beta$ E. Data are expressed as mean ( $\pm$ s.e.m.) total number of withdrawal signs observed at each time point or DH $\beta$ E dose. Rats are the same as those shown in Fig. 1a and b, respectively.

**Table 1 Somatic signs of spontaneous nicotine withdrawal**

| Withdrawal sign           | Group             | Hours post-pump removal |               |               |               |              |               |               |              |              |             |             |             |             |
|---------------------------|-------------------|-------------------------|---------------|---------------|---------------|--------------|---------------|---------------|--------------|--------------|-------------|-------------|-------------|-------------|
|                           |                   | 2.5                     | 4.5           | 6.5           | 8.5           | 12.5         | 16.5          | 24.5          | 36.5         | 48.5         | 80.5        | 104.5       | 128.5       | 152.5       |
| Abdominal constrictions   | Nicotine (s.e.m.) | 4.00* (1.18)            | 13.50* (3.11) | 13.75* (2.79) | 15.00* (4.30) | 7.13* (0.91) | 10.00* (2.52) | 12.38* (2.43) | 5.50* (1.62) | 6.50* (1.21) | 2.88 (1.04) | 1.75 (0.63) | 0.88 (0.32) | 0.88 (0.51) |
|                           | Saline (s.e.m.)   | 0.67 (0.46)             | 1.17 (0.52)   | 1.33 (0.83)   | 2.33 (0.92)   | 1.83 (0.66)  | 2.17 (1.11)   | 1.50 (0.84)   | 1.17 (0.44)  | 1.00 (0.57)  | 0.83 (0.44) | 0.83 (0.44) | 0.17 (0.72) | 1.33 (0.67) |
| Eye blinks                | Nicotine (s.e.m.) | 6.00 (1.33)             | 4.00 (0.93)   | 5.88 (1.08)   | 2.38 (0.53)   | 2.75 (1.06)  | 4.75* (0.88)  | 5.50* (1.34)  | 6.50* (0.97) | 6.38* (1.05) | 4.50 (1.01) | 4.38 (1.12) | 4.50 (0.90) | 4.25 (1.31) |
|                           | Saline (s.e.m.)   | 2.50 (1.16)             | 3.17 (0.59)   | 4.17 (1.00)   | 3.17 (1.28)   | 1.67 (0.54)  | 1.00 (0.57)   | 1.50 (0.47)   | 3.00 (0.49)  | 2.00 (0.57)  | 3.17 (0.59) | 1.83 (0.72) | 3.17 (0.96) | 2.50 (1.01) |
| Facial fasciculations     | Nicotine (s.e.m.) | 0.63 (0.53)             | 0.63 (0.53)   | 0.88 (0.32)   | 1.88 (0.91)   | 2.50 (0.97)  | 3.38 (1.27)   | 2.00 (0.93)   | 2.25 (1.23)  | 1.13 (0.47)  | 0.50 (0.54) | 0.38 (0.28) | 0.00 (0.00) | 0.13 (0.13) |
|                           | Saline (s.e.m.)   | 0.33 (0.37)             | 0.50 (0.55)   | 0.83 (0.44)   | 0.33 (0.37)   | 0.83 (0.34)  | 0.67 (0.54)   | 0.17 (0.18)   | 1.00 (0.40)  | 1.00 (0.57)  | 1.00 (0.49) | 0.17 (0.18) | 0.83 (0.59) | 0.83 (0.44) |
| Miscellaneous other signs | Nicotine (s.e.m.) | 0.63 (0.35)             | 1.38 (0.53)   | 0.88 (0.51)   | 0.63 (0.28)   | 1.13 (0.62)  | 1.75 (0.60)   | 2.25 (0.66)   | 1.88 (0.65)  | 1.75 (0.78)  | 1.38 (0.60) | 1.00 (0.64) | 0.88 (0.32) | 0.75 (0.53) |
|                           | Saline (s.e.m.)   | 0.67 (0.37)             | 1.33 (0.68)   | 0.50 (0.25)   | 0.33 (0.23)   | 1.33 (0.78)  | 1.33 (0.46)   | 2.33 (1.35)   | 0.83 (0.44)  | 2.50 (1.09)  | 1.83 (1.07) | 2.17 (0.87) | 2.50 (1.49) | 2.33 (0.88) |

Mean ( $\pm$ s.e.m.) number of somatic withdrawal signs for nicotine- ( $9.0 \text{ mg kg}^{-1} \text{ s}^{-1}$  for 7 d) ( $n = 8$ ) and saline-treated ( $n = 6$ ) rats during spontaneous nicotine withdrawal after removal of subcutaneous osmotic minipumps (rats are the same as those used for Figs 1a, 2a). Abdominal constrictions included gasps and writhes. Facial fasciculations included chewing, teeth chattering and cheek tremor. Miscellaneous other signs included ptosis, 'wet dog' shakes, piloerection, genital grooming and escapes.

\* Statistically significant differences ( $P < 0.05$ ) between saline and nicotine groups as determined by Tukey's Studentized Range Method after a statistically significant main effect in the ANOVA.

decreased function in brain reward systems during nicotine withdrawal, closely parallel the negative affective state experienced by smokers during the first few days of abstinence, when they report the strongest urges to smoke and the most severe nicotine withdrawal symptoms<sup>27–29</sup>. Our results indicate that elevations in ICSS reward thresholds observed in rats during nicotine abstinence may be a useful model of the affective aspects of nicotine withdrawal in humans. These results also demonstrate a profound perturbation within brain reward circuitries produced by chronic nicotine administration that may contribute to nicotine addiction. □

## Methods

**Experimental procedures.** Male Wistar rats (Charles River; 280–300 g at the start of the experiment) were anaesthetized with a halothane/oxygen mixture and were prepared with a stainless-steel bipolar electrode (Plastics One) in the posterior lateral hypothalamus (coordinates: anterior–posterior,  $-0.5 \text{ mm}$  from bregma; lateral,  $\pm 1.7 \text{ mm}$  from midline; ventral,  $-8.3 \text{ mm}$  from dura; incisor bar  $5.0 \text{ mm}$  above interaural line). The procedure used for measuring brain reward thresholds was a modification of a discrete-trial current-threshold procedure<sup>30</sup> described previously<sup>11</sup>. The duration of each ICSS session was  $\sim 30 \text{ min}$ . Stable baseline thresholds (defined as  $\pm 10\%$  of the mean on 5 consecutive days) were established for all rats before implantation of the minipumps.

For spontaneous nicotine withdrawal, osmotic minipumps (model 2ML1, Alza), filled with either saline or nicotine hydrogen tartrate dissolved in saline, were implanted under halothane/oxygen anaesthesia. The concentration of nicotine was adjusted to compensate for differences in body weight at the time of implantation ( $\sim 38 \text{ mg ml}^{-1}$  nicotine hydrogen tartrate), and the dose of nicotine hydrogen tartrate delivered was  $9.0 \text{ mg kg}^{-1} \text{ d}^{-1}$  ( $3.16 \text{ mg kg}^{-1} \text{ d}^{-1}$  nicotine base). Pumps were removed 7 days later under brief halothane anaesthesia. ICSS reward thresholds were determined at 2, 4, 6, 8, 12, 16, 24, 36, 48, 80, 104, 128 and 152 h, and once daily thereafter after pump removal.

For the DH $\beta$ E-precipitated withdrawal study, rats were prepared with saline- or nicotine-containing ( $9.0 \text{ mg kg}^{-1} \text{ d}^{-1}$  for 14 d;  $\sim 69 \text{ mg ml}^{-1}$  nicotine hydrogen tartrate) osmotic minipumps (model 2ML2, Alza) under the same conditions as those described for the spontaneous withdrawal study. On day 6 after minipump implantation, rats received the first injection of DH $\beta$ E (0.0, 0.5, 1.0, 2.0  $\text{mg kg}^{-1} \text{ s.c.}$ ,  $1.0 \text{ ml kg}^{-1}$ ) dissolved in 0.9% sterile saline, 15 min before the beginning of the ICSS test session, using a within-subjects Latin-square design. Animals were required to return to baseline ICSS threshold levels for at least one ICSS session before subsequent DH $\beta$ E or vehicle injections. On day 13 after minipump implantation, rats were prepared with a second 14-day minipump. After a baseline ICSS session, rats received the first DH $\beta$ E injection in a second within-subjects Latin-square design (DH $\beta$ E doses: 0.0, 2.0, 3.0,

$4.0 \text{ mg kg}^{-1} \text{ s.c.}$ ,  $1.0 \text{ ml kg}^{-1}$ , 15 min before ICSS testing). After completion of the second Latin-square and a subsequent baseline ICSS session, ICSS reward thresholds in all rats were assessed after a single injection of  $6.0 \text{ mg kg}^{-1}$  DH $\beta$ E.

During assessment of somatic withdrawal signs, each rat was placed into a cylindrical plastic observation chamber, and the frequency of abstinence symptoms was recorded for 10 min using an opiate-abstinence scale modified to score nicotine abstinence<sup>17</sup>. Experimenters were blind to the treatment of each rat during observation of somatic abstinence signs. Subjects were habituated to the observation room and chambers before baseline assessment. Rats in all groups were observed for somatic withdrawal signs on the day before minipump implantation, and on the sixth (precipitated withdrawal groups) or seventh (spontaneous withdrawal groups) day of nicotine administration. Somatic withdrawal signs in rats undergoing spontaneous withdrawal were observed immediately after each ICSS reward threshold session during the 7 days after minipump removal. Precipitated somatic withdrawal signs were observed immediately after the ICSS reward threshold session after administration of each dose of DH $\beta$ E,  $\sim 45 \text{ min}$  after the DH $\beta$ E or vehicle injection.

**Data analyses.** All data were analysed using within-subjects repeated-measures analyses of variance (ANOVAs), followed by Tukey's Studentized Range Method after observation of a statistically significant effect of treatment conditions in the ANOVA. In the spontaneous nicotine withdrawal experiment, 3 rats (2 nicotine- and 1 saline-treated) had one undefined ICSS reward threshold value at a single time point during the 13 repeated ICSS threshold determinations. Exclusion of all of the data from these 3 animals from all analyses did not alter the pattern of results or the statistical significance level at any time for any of the dependent measures. Therefore, to include the remainder of these animals' data in the statistical analyses, missing values (3 of 182 total threshold estimates;  $< 2\%$ ) were substituted with group means.

Received 30 December 1997; accepted 26 February 1998.

1. Malarcher, A. M., Chrismon, J. H., Giovino, G. A. & Erikson, M. P. Smoking-attributable mortality and years of potential life lost—United States, 1984. *Morb. Mort. Weekly Rep.* **46**, 444–451 (1997).
2. Stolerman, I. P. & Jarvis, M. J. The scientific case that nicotine is addictive. *Psychopharmacology* **117**, 2–10 (1995).
3. Risner, M. E. & Goldberg, S. R. A comparison of nicotine and cocaine self-administration in the dog: fixed-ratio and progressive-ratio schedules of intravenous drug infusion. *J. Pharmacol. Exp. Ther.* **224**, 319–326 (1983).
4. US Department of Health and Human Services. *The Health Consequences of Smoking: Nicotine Addiction: A Report of the U.S. Surgeon General*. (US Government Printing Office, Washington DC, 1988).
5. Hughes, J. R., Gust, S. W., Skoog, K., Keenan, R. M. & Fenwick, J. W. Symptoms of tobacco withdrawal: a replication and extension. *Arch. Gen. Psychiatry* **48**, 52–59 (1991).
6. Shiffman, et al. Progression from a smoking lapse to relapse: prediction from abstinence violation effects, nicotine dependence, and lapse characteristics. *J. Consult. Clin. Psychol.* **64**, 993–1002 (1996).
7. Markou, A., Kosten, T. R. & Koob, G. F. Neurobiological similarities in depression and drug dependence: a self medication hypothesis. *Neuropsychopharmacology* **18**, 135–174 (1998).
8. Damaj, M. I., Welch, S. P. & Martin, B. R. *In vivo* pharmacological effects of dihydro- $\beta$ -erythroidine, a nicotinic antagonist, in mice. *Psychopharmacology* **117**, 67–73 (1995).



9. Leith, N. J. & Barrett, R. J. Amphetamine and the reward system: evidence for tolerance and post-drug depression. *Psychopharmacologia* **46**, 19–25 (1976).
10. Markou, A. & Koob, G. F. Postcocaine anhedonia: an animal model of cocaine withdrawal. *Neuropsychopharmacology* **4**, 17–26 (1991).
11. Markou, A. & Koob, G. F. Construct validity of a self-stimulation threshold paradigm: effects of reward and performance manipulations. *Physiol. Behav.* **51**, 111–119 (1992).
12. Schulteis, G., Markou, A., Gold, L. H., Stinus, L. & Koob, G. F. Relative sensitivity to naloxone of multiple indices of opiate withdrawal: a quantitative dose–response analysis. *J. Pharmacol. Exp. Ther.* **271**, 1391–1398 (1994).
13. Schulteis, G., Markou, A., Cole, M. & Koob, G. F. Decreased brain reward produced by ethanol withdrawal. *Proc. Natl Acad. Sci. USA* **92**, 5880–5884 (1995).
14. Okamoto, M., Rosenberg, H. C. & Boisse, N. R. Withdrawal characteristics following chronic pentobarbital dosing. *Eur. J. Pharmacol.* **40**, 107–119 (1976).
15. Gellert, V. F. & Holzman, S. G. Development and maintenance of morphine tolerance and dependence in the rat by scheduled access to morphine drinking solutions. *J. Pharmacol. Exp. Ther.* **205**, 536–546 (1978).
16. Mendelson, J. H. & Mello, N. K. Medical progress: biological concomitants of alcoholism. *N. Engl. J. Med.* **301**, 912–921 (1979).
17. Malin, D. H. et al. Rodent model of nicotine abstinence syndrome. *Pharmacol. Biochem. Behav.* **43**, 779–784 (1992).
18. Hildebrand, B. E., Nomikos, G. G., Bondjers, C., Nisell, M. & Svensson, T. H. Behavioral manifestations of the nicotine abstinence syndrome in the rat: peripheral versus central mechanisms. *Psychopharmacology* **129**, 348–356 (1997).
19. Malin, D. H. et al. The nicotinic antagonist mecamylamine precipitates nicotine abstinence syndrome in the rat. *Psychopharmacology* **115**, 180–184 (1994).
20. Malin, D. H., Lake, J. R., Carter, V. A., Cunningham, J. S. & Wilson, O. B. Naloxone precipitates nicotine abstinence syndrome in the rat. *Psychopharmacology* **112**, 339–342 (1993).
21. Murrin, C. L., Ferrer, J. R., Wanyun, Z. & Haley, N. J. Nicotine administration to rats: methodological considerations. *Life Sci.* **40**, 1699–1708 (1987).
22. Benowitz, N. L. Pharmacological aspects of cigarette smoking and nicotine addiction. *N. Engl. J. Med.* **319**, 1318–1330 (1988).
23. Huston-Lyons, D. & Kornetsky, C. Effects of nicotine on the threshold for rewarding brain stimulation in rats. *Pharmacol. Biochem. Behav.* **41**, 755–759 (1992).
24. Baucó, P. & Wise, R. A. Potentiation of lateral hypothalamic and midline mesencephalic brain stimulation reinforcement by nicotine: examination of repeated treatment. *J. Pharmacol. Exp. Ther.* **271**, 294–301 (1994).
25. Nisell, M., Nomikos, G. G. & Svensson, T. H. Infusion of nicotine in the central tegmental area or the nucleus accumbens of the rat differentially affects accumbal dopamine release. *Pharmacol. Toxicol.* **75**, 348–352 (1994).
26. Pontieri, F. E., Tanda, G., Orzi, F. & Di Chiara, G. Effects of nicotine on the nucleus accumbens and similarity to those of addictive drugs. *Nature* **382**, 255–257 (1996).
27. Killen, J. D., Fortmann, S. P., Newman, B. & Varady, A. Prospective study of factors influencing the development of craving associated with smoking cessation. *Psychopharmacology* **105**, 191–196 (1991).
28. Hughes, J. R. et al. Smoking cessation among self-quitters. *Health Psychol.* **11**, 331–334 (1992).
29. Doherty, K., Kinnunen, T., Militello, F. S. & Garvey, A. J. Urges to smoke during the first month of abstinence: relationship to relapse and predictors. *Psychopharmacology* **119**, 171–178 (1995).
30. Kornetsky, C. & Esposito, R. U. Euphoric drugs: effects on the reward pathways of the brain. *Fed. Proc.* **38**, 2473–2476 (1979).

**Acknowledgements.** This research was funded by a Novartis grant to A.M., National Institute on Drug Abuse (NIDA, USA) grants to A.M. and G.F.K., and a HIDA Individual National Research Service Award to M.P.E.-J. This is publication number 11286-NP from The Scripps Research Institute. We thank M. Arends for assistance with manuscript preparation.

Correspondence and requests for materials should be addressed to A.M. (e-mail: amarkou@scripps.edu).

## Salmonella typhi uses CFTR to enter intestinal epithelial cells

Gerald B. Pier\*, Martha Grout\*, Tanweer Zaidi\*, Gloria Meluleni\*, Simone S. Mueschenborn\*, George Banting†, Rosemary Ratcliff‡, Martin J. Evans§ & William H. Colledge‡

\* Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

† Department of Biochemistry and Biotechnology and Biological Sciences Research Council, Molecular Recognition Centre, University of Bristol, Bristol BS8 1TD, UK

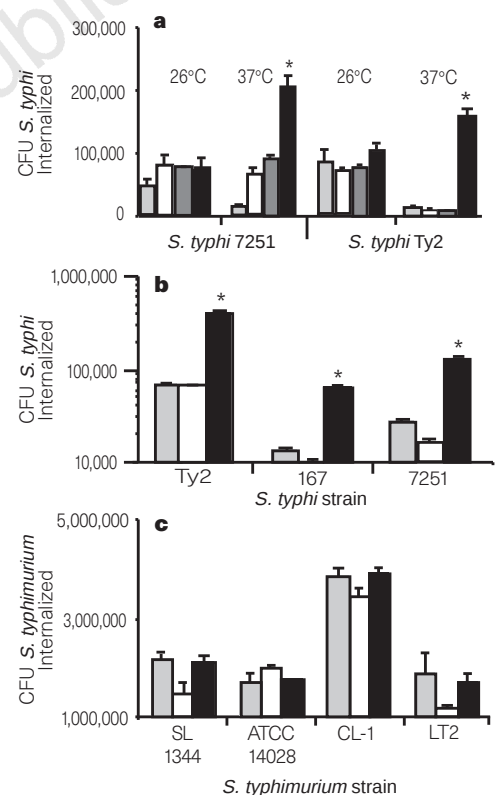
‡ Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

§ Wellcome/CRC Institute of Cancer and Developmental Biology and Department of Genetics, Tennis Court Road, University of Cambridge, Cambridge CB2 1QR, UK

**Homologous mutations of the cystic fibrosis transmembrane conductance regulator (CFTR) cause cystic fibrosis (CF). In the heterozygous state, increased resistance to infectious diseases may maintain mutant CFTR alleles at high levels in selected populations<sup>1</sup>. Here we investigate whether typhoid fever could be one such disease. The disease is initiated when *Salmonella typhi***

**enters gastrointestinal epithelial cells for submucosal translocation<sup>2</sup>. We found that *S. typhi*, but not the related murine pathogen *S. typhimurium*, uses CFTR for entry into epithelial cells. Cells expressing wild-type CFTR internalized more *S. typhi* than isogenic cells expressing the most common CFTR mutation, a phenylalanine deleted at residue 508 ( $\Delta$ 508). Monoclonal antibodies and synthetic peptides containing a sequence corresponding to the first predicted extracellular domain of CFTR inhibited uptake of *S. typhi*. Heterozygous  $\Delta$ 508 *Cftr* mice translocated 86% fewer *S. typhi* into the gastrointestinal submucosa than wild-type *Cftr* mice; no translocation occurred in  $\Delta$ 508 *Cftr* homozygous mice. The *Cftr* genotype had no effect on the translocation of *S. typhimurium*. Immunoelectron microscopy revealed that more CFTR bound to *S. typhi* in the submucosa of *Cftr* wild-type mice than in  $\Delta$ 508 heterozygous mice. We conclude that diminished levels of CFTR in heterozygotes may decrease susceptibility to typhoid fever.**

The entry of the *Salmonella enterica* subspecies *typhimurium* (*S. typhimurium*) into epithelial cells can be used as a model to define the bacterial and host factors involved in gastrointestinal (GI) translocation of the human-restricted pathogen *S. enterica* sub-

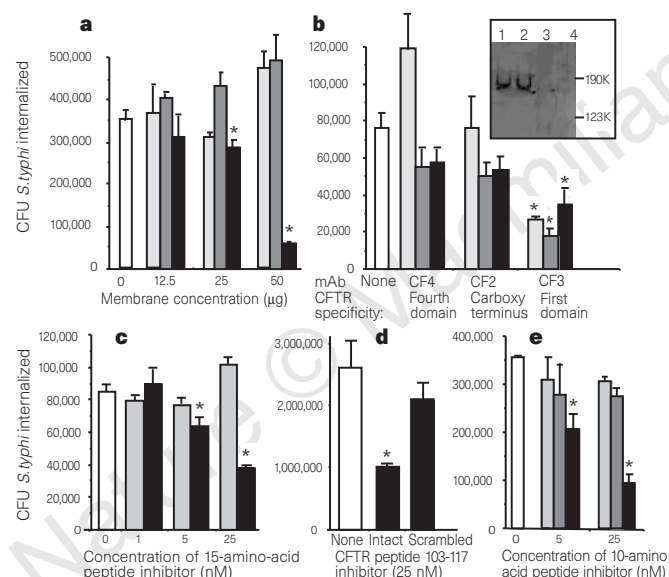


**Figure 1** Uptake of different *S. typhi* and *S. typhimurium* strains by epithelial cells expressing mutant  $\Delta$ 508 or wild-type *CFTR*. CFU, colony-forming units. Bars indicate means of 3–9 replicates and error bars the s.e.m. **a**, Ingestion of two *S. typhi* strains by airway epithelial cells differing in *CFTR* alleles and grown at either 26°C for 72 h or maintained at 37°C. Cells containing only  $\Delta$ 508 *CFTR* alleles include: CFT1 (parental line, stippled bar), CFT1-LC3 (vector control, white bars), and CFT1-F508 (3  $\Delta$ 508 *CFTR* alleles, cross-hatched bars). The CFT1-LCFSN cells (black bars) express wild-type human *CFTR*. Asterisks indicate significantly increased uptake of *S. typhi* compared with cells with  $\Delta$ 508 *CFTR* alleles only ( $P < 0.001$  by ANOVA and Fisher PLSD for all pairwise comparisons). **b, c**, Uptake of indicated *S. typhi* (**b**) or *S. typhimurium* (**c**) strains by murine C127 cells transfected with either no DNA (stippled bars), or DNA encoding either human  $\Delta$ 508 *CFTR* (white bars) or human wild-type *CFTR* (black bars) alleles. Asterisks indicate significantly increased uptake ( $P < 0.001$ , ANOVA and Fisher PLSD for pairwise comparisons). There were no significant differences ( $P > 0.3$ , ANOVA) for any of the *S. typhimurium* strains.

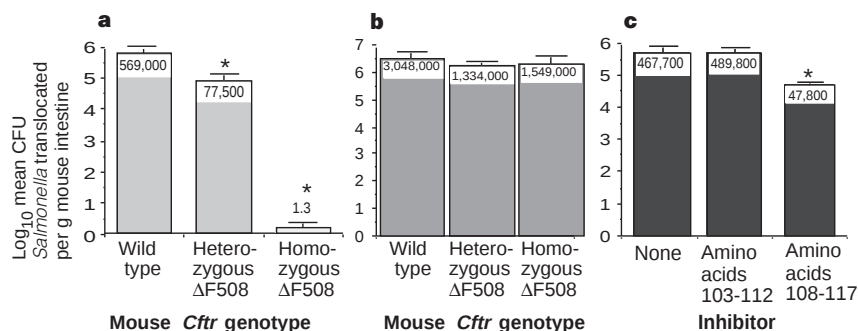
species *typhi* (*S. typhi*)<sup>2</sup>. Because uptake of these two pathogens by cultured epithelial cells is comparable<sup>3</sup>, differences in this process, such as receptor usage for internalization, may be involved in the different diseases caused by these pathogens in humans. We investigated whether CFTR could be a receptor for these pathogens by using a bacterial ingestion assay at 37°C in CFT1 epithelial cells, derived from a CF patient homozygous for the  $\Delta F508$  CFTR allele, and the isogenic cell line CFT1-LCFSN transfected with complementary DNA encoding wild-type CFTR<sup>4</sup>. We found that human epithelial cells expressing wild-type CFTR ingested significantly more *S. typhi* than cells expressing  $\Delta F508$  CFTR (Fig. 1a). Ingestion of *S. typhi* was the same as in wild-type cells when CFT1 cells were grown at 26°C, a temperature that allows for apical membrane expression of the  $\Delta F508$  CFTR protein at ~30% of wild-type levels<sup>5,6</sup> (Fig. 1a). We confirmed that the wild-type CFTR allele enhanced cell ingestion of *S. typhi* by measuring the uptake of three strains of *S. typhi* into C127 murine epithelioid cells expressing  $\Delta F508$  or wild-type CFTR alleles<sup>7</sup> (Fig. 1b). In contrast, there was no difference in uptake of *S. typhimurium* by C127 cells expressing either wild-type or  $\Delta F508$  CFTR (Fig. 1c). Results were identical with *S. typhimurium* and the CFT1 series of epithelial cells. Thus, *S. typhi* and *S. typhimurium* enter epithelial cells by different pathways, which may be one host component influencing the pathological

differences induced by these otherwise closely related organisms.

As *S. typhi* is a GI pathogen, we evaluated CFTR-mediated bacterial uptake by T84 human colon carcinoma cells, which express high levels of CFTR. *S. typhi* ingestion by these cells is inhibited by adding CFTR-specific reagents to the bacterial inoculum<sup>8</sup>. Plasma membranes isolated from murine C127 cells containing wild-type CFTR inhibited internalization of *S. typhi* by T84 cells, whereas membranes from cells expressing  $\Delta F508$  CFTR did not (Fig. 2a). A monoclonal antibody (CF3)<sup>9</sup> that recognizes the first extracellular domain of human CFTR also inhibited uptake of the bacteria, whereas antibodies against either the fourth predicted extracellular domain or a cytoplasmic domain of CFTR did not (Fig. 2b). Immunoblot analysis using the inhibitory CF3 detected a protein of relative molecular mass 170,000 ( $M_r$  170K) in immunoblot analysis of plasma membranes prepared from T84 cells and C127 cells expressing wild-type CFTR. In contrast, the same antibody failed to detect anything in plasma membranes from C127 cells expressing either no CFTR or  $\Delta F508$  CFTR (inset in Fig. 2b). To confirm the specificity of the CF3 antibody, we added different synthetic CFTR peptides to the antibody before using it to inhibit bacterial entry into cells. Only the cognate peptide of the first extracellular CFTR domain restored epithelial cell uptake of *S. typhi* by inhibiting CF3. A 15-amino-acid peptide derived from the first



**Figure 2** Inhibition of uptake of *S. typhi* Ty2 by T84 colonic epithelial cells with reagents specific to CFTR. **a**, Inhibition of uptake by membranes isolated from C127 murine epithelioid cells expressing no CFTR (stippled bars),  $\Delta F508$  CFTR (cross-hatched bars), or wild-type CFTR (black bars). **b**, Inhibition of uptake by monoclonal antibodies (mAbs) with the indicated specificity to CFTR domains<sup>9</sup> at antibody concentrations of 60  $\mu\text{g ml}^{-1}$  (stippled bars), 24  $\mu\text{g ml}^{-1}$  (cross-hatched bars), or 6  $\mu\text{g ml}^{-1}$  (black bars). Concentrations of mAb CF3 lower than that shown did not inhibit *S. typhi* uptake. Inset, binding of mAb CF3 to membranes extracted from (1) T84 cells; (2) C127 cells expressing wild-type human CFTR; (3) C127 cells expressing human  $\Delta F508$  CFTR; (4) untransfected C127 cells. Numbers on the right indicate relative molecular masses. **c**, **d**, Inhibition of uptake into T84 cell monolayers (**c**) or polarized cell cultures (**d**). In **c**, black bars represent the synthetic version of CFTR amino acids 103–117 and stippled bars the scrambled version of this peptide. **e**, Inhibition of uptake by a synthetic peptide corresponding to CFTR amino acids 108–117 (black bars), but not by a peptide of amino acids 103–112 (cross-hatched bars) or the scrambled version of peptide 103–117 (stippled bars). Bars indicate means, and error bars the s.e.m. of 6–9 replicates. Asterisks indicate differences from no inhibitor (white bar) significant at  $P \leq 0.01$  (ANOVA and Fisher PLSD for pairwise comparisons).



**Figure 3** *S. typhi* uses CFTR to translocate from the GI lumen to the submucosa. Bars indicate mean of the log<sub>10</sub> CFU of *Salmonella typhi* translocated; numbers indicate the antilog of the mean; error bars indicate the s.e.m. Asterisks indicate significant differences from wild-type mice ( $P < 0.001$ , ANOVA and Fisher PLSD for all pairwise comparisons). **a**, Decreased translocation of *S. typhi* Ty2 from the lumen of the GI tract of mice with the indicated genotype for murine CFTR. **b**, No

difference in translocation of *S. typhimurium* LT-2 among wild-type, heterozygous, and homozygous  $\Delta F508$  CFTR mice ( $P = 0.4$ , ANOVA). **c**, Inhibition of translocation of *S. typhi* Ty2 from the GI lumen of BALB/c mice infected with this bacterium, plus a synthetic peptide corresponding to the indicated amino acids in the first predicted extracellular domain of CFTR.

CFTR extracellular domain (CFTR residues 103–117) inhibited *S. typhi* ingestion by both monolayers and polarized cultures of T84 cells (Fig. 2c, d), whereas a scrambled synthetic peptide of these residues did not. CF3, but not the other CFTR-specific antibodies, also inhibited *S. typhi* uptake by polarized T84 cell cultures ( $P < 0.05$ , Fisher PLSD). *S. typhi* binding to CFTR was further narrowed down to a 10-amino-acid peptide (residues 108–117) that was sufficient to inhibit ingestion (Fig. 2e). Ingestion of *S. typhi* was no better in a cell line transfected with a cDNA encoding a truncated version of wild-type CFTR (C38 cells, which lack the first 150 amino acids of CFTR) than ingestion by the parental cell line (IB3 cells) derived from a CF patient carrying the *CFTR* alleles  $\Delta F508$  and W1243X (ref. 10) (data not shown). The pattern of inhibition of *S. typhi* ingestion was the same when LCFSN or HEP2 epithelial cells were used instead of T84 cells and with the same inhibitors (data not shown). Although all of these cell lines were efficient at ingesting *S. typhimurium*, we could not inhibit this uptake with CFTR-specific reagents (data not shown). These results indicate that CFTR is a major epithelial-cell receptor for internalization of *S. typhi*.

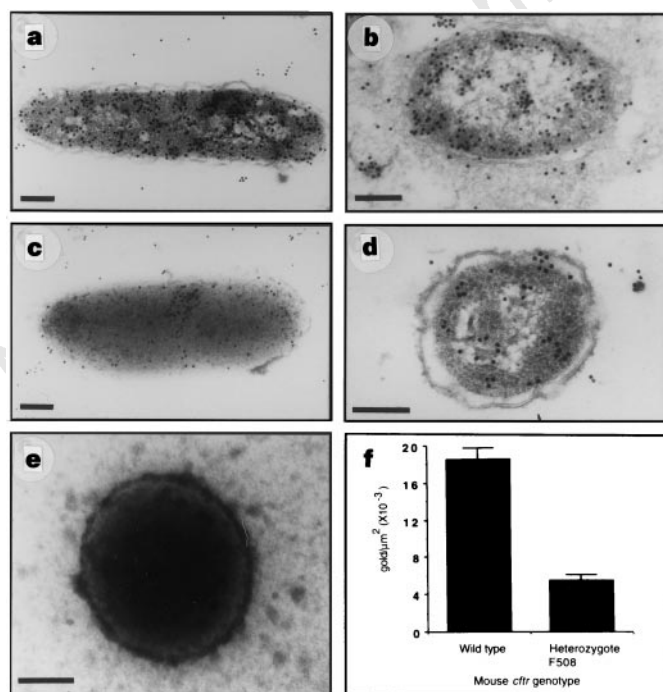
*S. typhi* does not cause systemic disease in mice when given orally, because *S. typhi* does not survive in murine macrophages<sup>11,12</sup>; however, *S. typhi* does enter macrophages and submucosal phagocytic cells in mouse GI epithelium in much the same way as does the virulent *S. typhimurium*<sup>13</sup>. This property makes CFTR-mediated translocation of *S. typhi* across the murine GI epithelium a suitable model for evaluating this step in *S. typhi* pathogenesis. We have shown previously that human CFTR-specific peptides and antibodies behave comparably in bacterial ingestion assays using either human or murine CFTR-expressing target cells<sup>8</sup>. To investigate the effect of CFTR expression on *S. typhi* submucosal translocation, we

compared sections of murine jejunum<sup>14</sup> from wild-type mice with littermate transgenic mice heterozygous or homozygous for the  $\Delta F508$  allele of murine *Cftr*<sup>15</sup>. In  $\Delta F508$  *Cftr* mice, the mutant CFTR does not leave the endoplasmic reticulum, fails to reach the plasma membrane, and is degraded<sup>16</sup>. In comparison with wild-type mice, translocation of *S. typhi* into the intestinal submucosa of  $\Delta F508$  *Cftr* heterozygotes was 86% less and in homozygotes it was almost zero (Fig. 3a). There was no significant difference among wild-type, heterozygous or homozygous  $\Delta F508$  *Cftr* mice in GI translocation of *S. typhimurium* (Fig. 3b), consistent with *in vitro* results<sup>13</sup>.

As the GI tracts of homozygous  $\Delta F508$  *Cftr* mice are abnormal<sup>15</sup>, we confirmed that CFTR mediates *S. typhi* translocation by injecting these bacteria into the GI lumen of BALB/c mice together with peptides corresponding to CFTR residues that either do or do not inhibit uptake into epithelial cells (Fig. 2d, e). Inclusion of the CFTR peptide 108–117 in the infectious inoculum reduced *S. typhi* translocation by 90% compared with animals given either the bacteria alone or a control peptide of CFTR amino acids 103–112 (Fig. 3c). From the correlation of the degree of *S. typhi* translocation with *Cftr* gene dosage and the reduction in translocation that is caused by inhibiting bacterial interaction with CFTR, we conclude that CFTR is an epithelial cell receptor used by *S. typhi* to enter the GI submucosa following oral intake.

We randomly selected electron micrographs of submucosal tissue from the jejunum of wild-type and heterozygous  $\Delta F508$  *Cftr* mice infected with *S. typhi* in order to quantify CFTR binding to bacteria after GI translocation. In wild-type mice, bacterial cells in the submucosa were heavily coated with CFTR after staining with monoclonal antibody CF8, which is specific for a cytosolic region of CFTR (Fig. 4a, b). Sections from heterozygous  $\Delta F508$  mice showed that CFTR binding to submucosal bacteria was significantly reduced (Fig. 4c, d), indicating that the amount of CFTR reaching the plasma membrane correlates with the extent of *S. typhi* translocation from the GI lumen. These differences were quantified by counting gold particles bound per square micron of the bacterial surface (see Methods): significantly more CFTR from infected wild-type mice attached to *S. typhi* compared with CFTR from infected heterozygotes (Fig. 4f). No *S. typhi*-infected cells were seen by electron microscopy in the submucosa of homozygous  $\Delta F508$  *Cftr* mice. Therefore, the smaller amount of CFTR that heterozygotes can mobilize to interact with *S. typhi* correlates with the reduced translocation of this organism across the GI epithelium. Although previous reports indicated that CF8 antibody cross reacts with unidentified antigens expressed by human cells in culture<sup>9</sup>, we did not find gold particles bound to GI tissues from  $\Delta F508$  *Cftr* homozygotes; however, in mice with at least one wild-type *Cftr* allele, the monoclonal antibody bound to cell membranes where bacteria were not present. Under our conditions, the antibody binds to tissues only from mice expressing wild-type CFTR and is therefore specific for CFTR.

These findings indicate that *S. typhi* and *S. typhimurium* use different epithelial-cell receptors for translocation into the GI submucosa, which might explain their different host ranges and disease symptoms, and that resistance to typhoid may have selected the  $\Delta F508$  allele *CFTR* to a high frequency in selected human populations. Maintenance of this allele at 4–5% in some caucasoid populations is best explained by over-dominance<sup>1</sup>, with resistance to a particular disease conferring an advantage for the heterozygote. Heterozygotes may also be resistant to the effects of cholera toxin<sup>17</sup>: for example,  $CF^{+/-}$  mice secrete less epithelial fluid following challenge with cholera toxin than do wild-type mice<sup>18</sup>. However, the acute secretory responses to a variety of agonists, including cholera toxin, are reported by others to be similar in heterozygous and wild-type CFTR mice<sup>19</sup>. Cholera has been endemic in India for centuries<sup>20</sup> but did not appear in Europe until 1832 (refs 1, 20, 21), so cholera itself is unlikely to have selected for mutant alleles of *CFTR*. Pathogens that cause diarrhoea by producing other, similar



**Figure 4** Immunoelectron micrographs of *S. typhi* Ty2 in the submucosal space of the GI tract of mice reacted with a monoclonal antibody to CFTR. **a, b**, *S. typhi* in the submucosa of a mouse homozygous for wild-type CFTR. **c, d**, *S. typhi* in the submucosa of a mouse heterozygous for wild-type and  $\Delta F508$  CFTR. **e**, Control, with the primary antibody omitted. **f**, The number of gold particles per square micron bound to *S. typhi* Ty2 in the submucosa of mice homozygous for wild-type *Cftr* ( $n = 24$  bacterial cells counted) or heterozygous for  $\Delta F508$  *Cftr* ( $n = 55$  bacterial cells counted). Bars indicate means and error bars s.d. the difference in gold particles bound was significant at  $P < 0.001$  (*t*-test).

toxins may have selected for  $\Delta F508$  CFTR. Our results show that resistance to typhoid fever could serve as the selective factor for the heterozygote advantage conferred by the  $\Delta F508$  CFTR allele.  $\square$

## Methods

**Bacterial ingestion assays.** Wild-type or  $\Delta F508$ -expressing CFT1 human cell lines and C127 murine epithelioid cells were grown for 72 h at 26 °C or 24 h at 37 °C<sup>4,7</sup> before bacterial-uptake assay<sup>6,22</sup>. Bacterial strains were either clinical isolates or from the American Type Culture Collection. The bacterial inoculum was prepared by growing *S. typhi* overnight at 37 °C in L-broth and dilutions were made in supplemented tissue-culture medium. About  $1-3 \times 10^6$  CFU in 100  $\mu$ l was added to each monolayer of  $10^5$  epithelial cells and  $\sim 8 \times 10^6$  CFU were added to polarized cultures of T84 cells. Bacteria were allowed to invade the epithelial cells for 3–4 h at 37 °C, after which non-adherent bacteria were removed by washing. The number of ingested bacteria was quantified as described<sup>6,22</sup>.

**Inhibition of bacterial uptake.** T84 cells (American Type Culture Collection CCL-248) were grown in a 1:1 mixture of Ham's F-12 and Dulbecco's modified Eagle's medium with 10% fetal bovine serum. They were seeded either into 96-well tissue-culture plates at  $10^5$  cells per well or onto Transwell polycarbonate filters (0.4- $\mu$ m pores) and grown for 6 d to form a polarized monolayer. Polarized cells were used when the electrical resistance was  $\geq 2,000 \Omega \text{ cm}^{-2}$  and when  $<5\%$  of the c.p.m. of  $^3\text{H}$ -labelled mannitol added to the apical side of the cells was recovered 3 h later on the basal side<sup>23</sup>. Plasma membranes were prepared from C127 and T84 cells as described<sup>24</sup>. Membranes from C127 cells were suspended in 150 mM NaCl, 50 mM Tris (pH 7.5), 1 mM EDTA, and added at the indicated concentrations to suspensions of *S. typhi*. Monoclonal antibodies against CFTR<sup>9</sup> at stock concentrations of 600  $\mu\text{g ml}^{-1}$  were diluted from 1:10 to 1:1500 and added to bacterial inocula. Specificity of the monoclonal antibodies for CFTR in T84 and C127 cell membranes<sup>24</sup> was tested by immunoblotting<sup>9</sup>. Synthetic peptides mimicking the human CFTR first predicted extracellular domain were dissolved in F12 medium, and added at the indicated concentrations to bacterial inocula. These mixtures were incubated for 30 min at room temperature, then 100  $\mu$ l was added to  $10^5$  T84 cells growing in monolayers or to polarized T84 cells grown on filter supports. Internalized *S. typhi* were quantified as described<sup>22</sup>.

**Bacterial translocation.** Six-to-eight-week-old  $\Delta F508$  Cfr transgenic mice<sup>15</sup> were given water and a liquid elemental diet of Peptamen (Clintec Nutrition), each containing 1 mg gentamicin and 1 mg streptomycin per ml. Four days later, antibiotics were removed from the food and water and fecal pellets were checked for the presence of aerobic Gram-negative bacteria on MacConkey and *Salmonella*–*Shigella* agar plates: there was no growth on these media. Twenty-four hours later mice were killed and jejunal sections were cut into  $\sim 2$  cm lengths, closed at each end and injected with  $10^7$  CFU of *S. typhi* Ty2 or *S. typhimurium* LT2. For controls, colon sections were prepared similarly. After 2 h at 37 °C in 5% CO<sub>2</sub>, 200  $\mu\text{g}$  gentamicin in 50  $\mu$ l was injected into the lumen, and the tissue section was immersed in a solution of 400  $\mu\text{g}$  gentamicin per ml. After 1 h at 37 °C, the lumen was exposed, the tissue was weighed and then washed in 50 ml of F12 tissue-culture medium to remove gentamicin, homogenized in 0.5% Triton X-100 and then diluted for counting of bacterial cells on *Salmonella*–*Shigella* agar. Additional controls, together with the colon sections, included jejunal sections not treated with gentamicin (total growth) and jejunal sections first homogenized and then immersed in gentamicin to ensure that the antibiotic was sufficient to kill extracellular bacteria (total kill). In all cases,  $>99.9\%$  of the bacteria were killed if antibiotic was added after tissue homogenization; no *Salmonella* surviving gentamicin treatment were obtained from control colon sections. Statistical comparisons were made using the Kruskal–Wallis non-parametric ANOVA and Dunn procedure for pairwise comparison.

Inhibition by CFTR peptide of *S. typhi* translocation was tested on jejunal sections from BALB/c mice. The bacterial inoculum was delivered together with 25 nM human CFTR peptide (residues 103–112), which differs from mouse CFTR at position 112 (aspartic to glutamic acid), or another human CFTR peptide containing residues 108–117 (glutamic acid substituted by valine at position 115 in human CFTR).

**Electron microscopy.** For antibody-binding studies, tissues in 2% glutaraldehyde were first sectioned for electron microscopy. Nickel-coated grids with

tissue sections were placed on a 5- $\mu$ l drop of a 1:10 dilution of antibody CF8, which is specific for a 16-amino-acid region of both human and murine CFTR (which differ by a single conservative serine-to-alanine substitution at the third position of the peptide) located in the cytoplasmic domain connecting the fifth and sixth transmembrane domains and encompassing human CFTR residues 1,035–1,050 (ref. 9). Controls had no antibody. After 1 h at 25 °C, grids were washed and placed on a 5- $\mu$ l drop of rabbit antibodies raised against mouse IgM. After 30 min at room temperature, grids were washed and placed on a drop of protein A labelled with 20-nm gold particles, then incubated for 30 min. After three washes, sections were stained with osmium tetroxide and processed for electron microscopy.

Cells infected with *S. typhi* were identified on grids at a magnification (4,000–5,000 $\times$ ) too low to resolve individual gold particles, so the magnification was increased to 25,000–30,000. Two observers independently measured the bacterial surface area on the micrographs, using formulae for a circle or an ellipse, and counted the gold particles on the bacterial surface. There were no significant differences ( $P = 0.8$ ,  $t$ -test) between the two observers' results, which were averaged to get the number of gold particles bound per  $\mu\text{m}^2$  on *S. typhi* in the submucosa of wild-type Cfr mice or heterozygous  $\Delta F508$  mice. Results were compared by  $t$ -test.

Received 30 December 1997; accepted 16 February 1998

- Bertranpetit, J. & Calafell, F. in *Variation in the Human Genome* (eds Chadwick, D. & Cardew, G.) 97–114 (Wiley, Chichester, 1996).
- Jones, B. D. & Falkow, S. Salmonellosis: host immune responses and bacterial virulence determinants. *Annu. Rev. Immunol.* **14**, 533–561 (1996).
- Mills, S. D. & Finlay, B. B. Comparison of *Salmonella typhi* and *Salmonella typhimurium* invasion, intracellular growth and localization in cultured human epithelial cells. *Microbiol. Path.* **17**, 409–423 (1994).
- Olsen, J. C. *et al.* Correction of the apical membrane chloride permeability defect in polarized cystic fibrosis airway epithelia following retroviral-mediated gene transfer. *Hum. Gene Ther.* **3**, 253–266 (1992).
- Denning, G. M. *et al.* Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive. *Nature* **358**, 761–764 (1992).
- Pier, G. B. *et al.* Role of mutant CFTR in hypersusceptibility of cystic fibrosis patients to lung infections. *Science* **271**, 64–67 (1996).
- Cheng, S. H. *et al.* Functional activation of the cystic fibrosis trafficking mutant  $\Delta F508$ -CFTR by overexpression. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **12**, L615–L624 (1995).
- Pier, G. B., Grout, M. & Zaidi, T. S. Cystic fibrosis transmembrane conductance regulator is an epithelial cell receptor for clearance of *Pseudomonas aeruginosa* from the lung. *Proc. Natl Acad. Sci. USA* **94**, 12088–12093 (1997).
- Walker, J., Watson, J., Holmes, C., Edelman, A. & Banting, G. Production and characterisation of monoclonal and polyclonal antibodies to different regions of the cystic fibrosis transmembrane conductance regulator (CFTR): detection of immunologically related proteins. *J. Cell Sci.* **108**, 2433–2444 (1995).
- Flotte, T. R. *et al.* Stable *in vivo* expression of the cystic fibrosis transmembrane conductance regulator with an adeno-associated virus vector. *Proc. Natl Acad. Sci. USA* **90**, 10613–10617 (1993).
- Alpuche-Aranda, C. M., Berthiaume, E. P., Mock, B., Swanson, J. A. & Miller, S. I. Spacious phagosome formation within mouse macrophages correlates with *Salmonella* serotype pathogenicity and host susceptibility. *Infect. Immun.* **63**, 4456–4462 (1995).
- Ishibashi, Y. & Arai, T. A possible mechanism for host-specific pathogenesis of *Salmonella* serovars. *Microbiol. Path.* **21**, 435–446 (1996).
- Pascope, L. *et al.* Host restriction phenotypes of *Salmonella typhi* and *Salmonella gallinarum*. *Infect. Immun.* **63**, 4329–4335 (1995).
- Jones, B. D., Ghorri, N. & Falkow, S. *Salmonella typhimurium* initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches. *J. Exp. Med.* **180**, 15–23 (1994).
- Colledge, W. H. *et al.* Generation and characterization of a  $\Delta F508$  cystic fibrosis mouse model. *Nature Genet.* **10**, 445–452 (1995).
- Tsui, L. C. The cystic fibrosis transmembrane conductance regulator gene. *Am. J. Respir. Crit. Care Med.* **151**, S47–S53 (1995).
- Romeo, G., Devoto, M. & Galietta, L. J. V. Why is the cystic fibrosis gene so frequent? *Human Genet.* **84**, 1–5 (1989).
- Gabriel, S. E., Brigman, K. N., Koller, B. H., Boucher, R. C. & Stutts, M. J. Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science* **266**, 107–109 (1994).
- Cuthbert, A. W., Halstead, J., Ratcliff, R., Colledge, W. H. & Evans, M. J. The genetic advantage hypothesis in cystic fibrosis heterozygotes: a murine study. *J. Physiol. (Lond)* **482**, 449–454 (1995).
- Pollitzer, R. *Cholera* (World Health Organization, Geneva, 1959).
- Van Heyningen, W. E. & Seal, J. R. *Cholera: The American Scientific Encounter, 1947–1980*. 1–343 (Westview, Boulder, Colorado, 1983).
- Feiszig, S. M. J., Zaidi, T. S. & Pier, G. B. *Pseudomonas aeruginosa* invasion of and multiplication within corneal epithelial cells *in vitro*. *Infect. Immun.* **63**, 4072–4077 (1995).
- Beatty, W. L. & Sansonetti, P. J. Role of lipopolysaccharide in signaling to subepithelial polymorphonuclear leukocytes. *Infect. Immun.* **65**, 4395–4404 (1997).
- O'Riordan, C. R. *et al.* Purification and characterization of recombinant cystic fibrosis transmembrane conductance regulator from Chinese hamster ovary and insect cells. *J. Biol. Chem.* **270**, 17033–17043 (1995).
- Rosner, B. in *Fundamentals of Biostatistics* 498–503 (Duxbury, Boston, Massachusetts, 1990).

**Acknowledgements.** We thank J. Olsen, J. Yankaskas and L. Johnson for CFT1 cells, A. Smith and colleagues (Genzyme) for C127 cell lines, P. Zeitlin for IB3 and C38 cells, C. Lee for advice on *Salmonella* uptake assays, and C. Lee and D. Kasper for reading the manuscript. This work was supported by the NIH, the Cystic Fibrosis Foundation and the Cystic Fibrosis Trust.

Correspondence and requests for materials should be addressed by G.B.P. (e-mail: gpier@channing.harvard.edu).



# Requirement of ErbB2 for signalling by interleukin-6 in prostate carcinoma cells

Yun Qiu\*, Lakshmeswari Ravi\* & Hsing-Jien Kung\*†

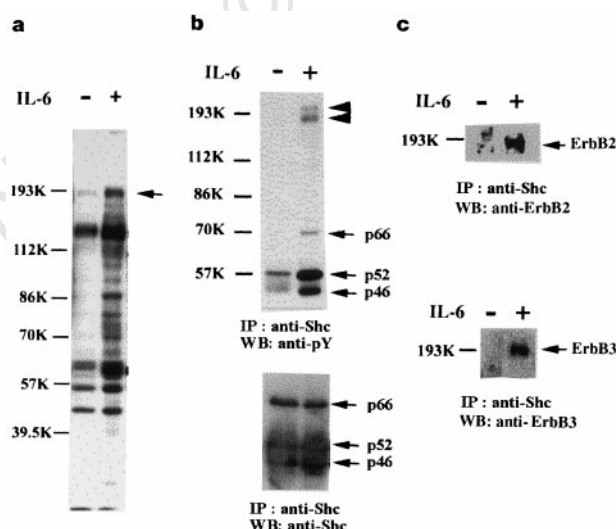
\* Department of Molecular Biology and Microbiology, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA

† Molecular and Genomic Medicine Division, National Health Research Institutes, 128 Yen-Chiu Yuan Road, Sec2, Taipei, Taiwan

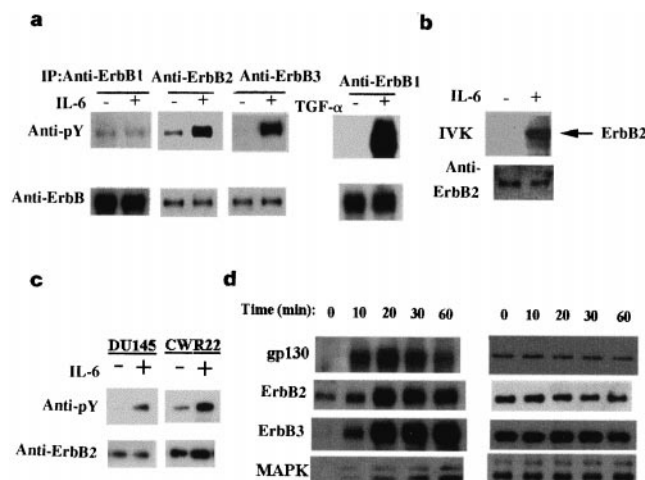
Interleukin-6 (IL-6) is a cytokine that was initially recognized as a regulator of immune and inflammatory responses<sup>1</sup>, but it also regulates the growth of many tumour cells, including prostate carcinoma<sup>2-4</sup>. Overexpression of the growth-factor receptors ErbB2/neu and ErbB3 has been implicated in the neoplastic transformation of prostate carcinoma<sup>5-7</sup>. Here we show that treatment of the prostate cancer cell line LNCaP with IL-6 induces tyrosine phosphorylation of ErbB2 and ErbB3, but not ErbB1/EGFR. We also show that ErbB2 forms a complex with the gp130 subunit of the IL-6 receptor in an IL-6-dependent manner. This association is important because the inhibition of ErbB2 activity results in abrogation of IL-6-induced MAPK activation. Thus ErbB2 is a critical component of IL-6 signalling through the MAP kinase pathway. These data show how a cytokine receptor can diversify its signalling pathways by engaging with a growth-factor receptor kinase.

In an attempt to study IL-6 signal transduction in prostate epithelial cells, we analysed the tyrosine phosphorylation pattern of IL-6-treated LNCaP cells. We found increased tyrosine phosphorylation of cellular proteins (Fig. 1a); in particular there are newly phosphorylated bands corresponding to receptor tyrosine kinases of relative molecular mass 190,000 ( $M_r$  190K). To define

specific signalling molecules activated by IL-6, we examined the tyrosine phosphorylation of Shc (Fig. 1b) and found that all three isoforms were phosphorylated upon treatment with IL-6. Two tyrosine-phosphorylated proteins around 190K are co-precipitated with Shc in IL-6-treated cells. These two proteins can be recognized by ErbB2 and ErbB3 antibodies, respectively (Fig. 1c), suggesting that ErbB2 and ErbB3 are tyrosine phosphorylated and Shc is recruited to associate with them in IL-6-treated LNCaP. We then examined the effect of IL-6 on the tyrosine phosphorylation of ErbB1, ErbB2 and ErbB3, which are expressed in LNCaP<sup>7</sup>. Treatment with IL-6 results in robust activation of ErbB2 and ErbB3, but not ErbB1 (Fig. 2a). Treatment with TGF- $\alpha$ , on the other hand, activates ErbB1, indicating that ErbB1 is functional in this cell type. This finding indicates the specificity of IL-6 activation. An *in vitro* kinase assay (Fig. 2b) indicates that the heightened tyrosine phosphorylation of ErbB2 results from the increase of the intrinsic ErbB2 kinase activity. We also observed that ErbB2 was activated by IL-6 in two other prostate carcinoma cell lines, DU145 and CWR22 (Fig. 2c), indicating that this phenomenon is not peculiar to LNCaP. We analysed the activation kinetics of ErbB2 and ErbB3 by IL-6 in LNCaP (Fig. 2d), using gp130, the  $\beta$ -subunit of the IL-6 receptor, as a control. Tyrosine-phosphorylated gp130 can be detected within 10 min of IL-6 treatment. The activation of ErbB2 and ErbB3 closely follow, and may be coincident with, that of gp130. The activation of MAP kinase follows a similar time course. The rapid kinetics of ErbB2 and ErbB3 activation makes it unlikely that it results from transcriptional activation of NDF/heregulin, which is the ligand for ErbB2 and ErbB3 (refs 7-9). Indeed, NDF is undetectable in LNCaP either before or after IL-6 treatment (data not shown). These data point to possible direct activation of ErbB2 and ErbB3 by the activated gp130. This possibility is supported by the co-precipitation of ErbB2 with gp130 in response to IL-6 stimulation (Fig. 3a). A tyrosine-phosphorylated 190K component (top panel), identified as ErbB2 by an anti-ErbB2 blot (middle panel), is associated with gp130 immunoprecipitates in an IL-6-dependent manner. Conversely, gp130 is detected in the ErbB2 immunoprecipitates from IL-6-treated LNCaP cells (Fig. 3b); an anti-haemagglutinin antibody was



**Figure 1** Association of ErbB2 and ErbB3 with Shc in response to IL-6 in LNCaP cells. **a**, Anti-phosphotyrosine western blot of the total extracts from LNCaP cells. The arrow indicates the 190K tyrosine-phosphorylated protein(s).  $M_r$  values (k) are shown on the left. **b**, Anti-phosphotyrosine (top) and anti-Shc (bottom) western blots of anti-Shc immunoprecipitates. The arrowheads indicate the two tyrosyl phosphorylated bands around 190K. **c**, As **b**, except probed by anti-ErbB2 (top) and anti-ErbB3 (bottom) antibodies (NeoMarkers). IP, immunoprecipitate; WB, western blot.



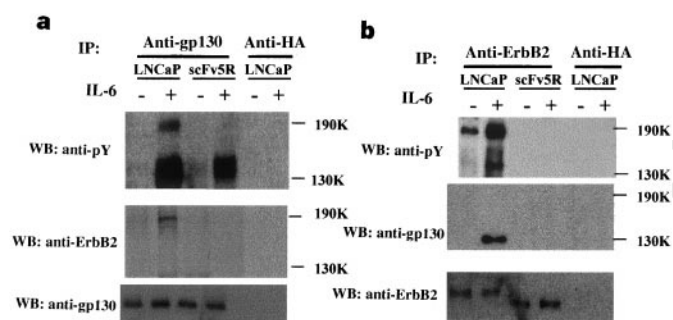
**Figure 2** IL-6 stimulates the tyrosine phosphorylation of ErbB2 with rapid kinetics. **a**, Anti-phosphotyrosine (top) and anti-ErbB (bottom) western blots of immunoprecipitates of ErbB1, ErbB2 and ErbB3, respectively, from LNCaP. Cells treated with TGF- $\alpha$  were used as an ErbB1-positive control. **b**, *In vitro* kinase (IVK) assay (top) and anti-ErbB2 western blot of anti-ErbB2 immunoprecipitates. **c**, Anti-phosphotyrosine (top) and anti-ErbB2 (bottom) western blots of anti-ErbB2 immunoprecipitates from DU145 and CWR22 cell lines. **d**, Anti-phosphotyrosine western blots of immunoprecipitates of gp130, ErbB2, ErbB3 and MAP kinase (left). Right, as left, except using antibodies to individual signalling molecules.

used as a negative control. Taken together, these data show that IL-6 induces the formation of a complex between gp130 and ErbB2. The activated gp130 dimer can potentially recruit two ErbB2 molecules, leading to their transphosphorylation.

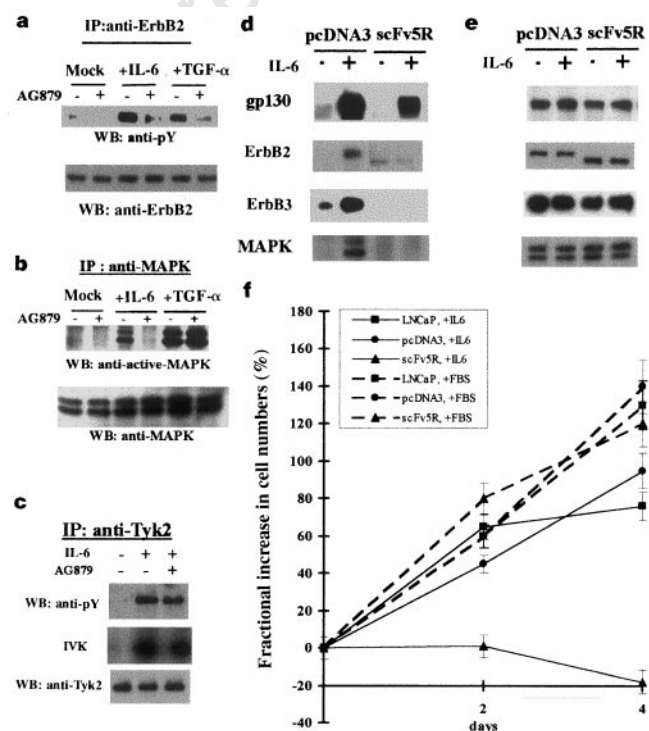
As described above, IL-6 activation of MAP kinase parallels that of ErbB2. Because ErbB2 is a strong activator of MAP kinase<sup>10</sup>, we investigated whether MAPK activation by IL-6 requires ErbB2. We used two approaches to make ErbB2 inactive. First, the tryphostin AG879 (ref. 11), a specific inhibitor of ErbB2 kinase activity, was tested for its effect on IL-6-stimulated MAP kinase activity. Pre-treatment with AG879 significantly reduced the phosphorylation of ErbB2 induced by IL-6 (Fig. 4a). Concomitant with AG879 treatment, IL-6-induced MAP kinase activity was also lost. However, this treatment had little effect on either the activation of MAP kinase stimulated by TGF- $\alpha$  (Fig. 4b) or IL-6-induced Tyk2 activity (Fig. 4c). These data suggest that the intrinsic tyrosine kinase activity of ErbB2 is required for the activation of MAPK by IL-6. The second strategy used to inactivate ErbB2 activity used a single-chain antibody against ErbB2, scFv5R (ref. 12), which traps ErbB2 in the endoplasmic reticulum. The LNCaP cells stably expressing the

gene encoding scFv5R were obtained as described in the Methods. The ErbB2 in scFv5R cells was expressed at a similar level to the control (Fig. 4e), except that it migrated faster, presumably as a result of underglycosylation<sup>12</sup>. Although the expression of this gene in LNCaP had little effect on the tyrosine phosphorylation of gp130, the activity of ErbB2 was completely abolished (Fig. 4d), and the association between gp130 and ErbB2 was no longer detectable (Fig. 3). In scFv5R cells, no MAP kinase activation could be detected following IL-6 treatment (Fig. 4d). These two experiments provide strong evidence that in LNCaP cells activation of MAP kinase by IL-6 is mediated through ErbB2. As another test of the function of ErbB2, we monitored the phosphorylation of ErbB3, which is a kinase-impaired receptor, the activation of which is thought to rely on transphosphorylation by ErbB2 through heterodimerization<sup>13,14</sup>. ErbB3 was properly expressed and processed, but there was no phosphorylation detected in scFv5R cells (Fig. 4d, e). Thus, ErbB2 is required to activate both ErbB3 and MAP kinase in IL-6-treated LNCaP. Further, ErbB2 is required for IL-6-stimulated growth of LNCaP cells<sup>4</sup> (Fig. 4, solid lines). IL-6 stimulated the growth of both LNCaP and pcDNA3 cells but not scFv5R cells; on the other hand, fetal bovine serum-stimulated growth (Fig. 4f, dotted lines) was not affected in scFv5R cells, indicating that the transfected gene does not have a general inhibitory effect.

Here we have shown that ErbB2 is a mediator of IL-6 responses in a prostate carcinoma cell line, LNCaP. Both MAP kinase activation and IL-6 induced cell growth require functional ErbB2. This is different from the case in haematopoietic cells<sup>15-18</sup>, which often lack ErbB kinases. IL-6 receptors can thus recruit different signal transducers in different cells to generate varied responses. IL-6 is known to induce the growth of a variety of prostate carcinoma cells<sup>4,19</sup>. Similarly, overexpression of ErbB2 and ErbB3 has been implicated in prostate carcinoma growth and progression<sup>5,6</sup>. The pleiotropic effects and potency of IL-6 on prostate cancer cell growth may be modulated by the presence and level of the various ErbB kinases. We also showed that the activated gp130 and ErbB2 receptors interact directly with each other. These interactions result in the increase of ErbB2 intrinsic kinase activity, which is required for autophosphorylation and the activation of ErbB3 and MAPK. We have explored the possible involvement of other non-receptor



**Figure 3** Co-precipitation of ErbB2 and gp130 in response to IL-6 stimulation in LNCaP. **a**, Anti-phosphotyrosine (top), anti-ErbB2 (middle) and anti-gp130 (bottom) western blots of anti-gp130 immunoprecipitates. Anti-haemagglutinin (HA) immunoprecipitates were used as a negative control. **b**, As **a**, except that the anti-ErbB2 immunoprecipitates were used. *M<sub>r</sub>* values (k) are shown.



**Figure 4** AG879 and scFv5R block the MAP kinase activation and cell growth induced by IL-6. **a-c**, LNCaP cells were pre-treated with 10  $\mu$ M AG879 for 15 min, followed by treatment of IL-6 for 30 min or TGF- $\alpha$  for 10 min as indicated. **a**, Anti-phosphotyrosine (top) and anti-ErbB2 (bottom) western blots of anti-ErbB2 immunoprecipitates. **b**, Anti-active-MAPK (top) and anti-MAPK (bottom) western blots of anti-MAPK immunoprecipitates. **c**, Anti-phosphotyrosine (top) and anti-Tyk2 (bottom) western blots of anti-Tyk2 immunoprecipitates. Middle, *in vitro* kinase assay of Tyk2. **d**, Anti-phosphotyrosine western blots of immunoprecipitates of gp130, ErbB2, ErbB3 and MAPK from LNCaP and scFv5R cells. **e**, As **d**, but probed with antibodies to gp130, ErbB2, ErbB3 and MAPK. **f**, Cell proliferation assays of LNCaP, pcDNA3-transfected cells and scFv5R-transfected cells in response to IL-6 (solid lines) or 1% fetal bovine serum (dotted lines). Data are expressed as the fractional increase in cell numbers relative to the serum-starved controls of each cell line at a given day. Error bars shown standard deviation of the duplicate samples.

tyrosine kinases, including the Jak family kinases, and have failed to identify any such kinase in the complex. Our current model is that the activated gp130 receptor dimers recruit ErbB2 molecules to the complex, leading to ErbB2 clustering and kinase activation. It is difficult to rule out the transient participation of other molecule(s) in this process. Once ErbB2 is activated, the Shc/MAP kinase pathway will be activated, as well as ErbB3. The native ligands of the ErbB-family kinases are mostly peptide growth factors containing EGF-like repeats<sup>20</sup>. Recently, however, ErbB1/EGFR kinase was found to mediate signals of ultraviolet irradiation<sup>21</sup>, G proteins<sup>22</sup> and growth hormone<sup>23</sup>. Our data extend these observations and identify ErbB2 and ErbB3 as signal mediators of a cytokine. □

## Methods

**Cell culture, transfection, selections and cell proliferation.** LNCaP, DU145 and CWR22 cells were maintained as described<sup>7</sup>. The cells were serum starved for 24 h followed by treatment with 200 ng ml<sup>-1</sup> IL-6 (UBI) for 30 min or the times indicated. The cDNA of scFv5R was subcloned into a mammalian expression vector, pcDNA3 (Invitrogen). LNCaP cells were transfected with this scFv5R construct by using LipofectAmine(Gibco BRL) following the manufacturer's instructions. The stable expression clones were obtained by selection for G418 resistance (600 µg ml<sup>-1</sup>) and further confirmed by the mobility shift of ErbB2 as described<sup>12</sup>. The cell proliferation assays (WST assays) were performed as described<sup>4</sup>.

**Immunoprecipitation, western blotting and *in vitro* kinase assays.** The immunoprecipitation and western blotting were performed following procedures described previously<sup>7</sup>. The autophosphorylation of ErbB2 and Tyk2 was assayed as described<sup>14,24</sup>. Briefly, the ErbB2 or Tyk2 immunoprecipitates were incubated in 20 mM HEPES, pH 7.4 with 10 mM MnCl<sub>2</sub> and 10 µCi [ $\gamma$ -<sup>32</sup>P]ATP for 10 min at room temperature. Reactions were stopped by adding an equal volume of 2 × SDS loading buffer and proteins resolved by SDS-PAGE, followed by autoradiography.

Received 23 December 1997; accepted 19 February 1998.

- Hirano, T. The biology of interleukin-6. *Chem. Immunol.* **51**, 153–180 (1992).
- Siegmund, M. J., Yamazaki, H. & Pastan, I. Interleukin 6 receptor mRNA in prostate carcinomas and benign prostate hyperplasia. *J. Urol.* **151**, 1396–1399 (1994).
- Hutchins, D. & Steel, C. M. Regulation of ICAM-1 (CD54) expression in human breast cancer cell lines by interleukin 6 and fibroblast-derived factors. *Int. J. Cancer* **58**, 80–84 (1994).
- Okamoto, M., Lee, C. & Oyasu, R. Interleukin-6 as a paracrine and autocrine growth factor in human prostatic carcinoma cells *in vitro*. *Cancer Res.* **57**, 141–146 (1997).
- Myers, R. B., Srivastava, S., Oelschlagel, D. K. & Grizzle, W. E. Expression of p160erbB-3 and p185erbB-2 in prostatic intraepithelial neoplasia and prostatic adenocarcinoma. *J. Nat. Cancer Inst.* **86**, 1140–1145 (1994).
- Zhou, H. E. *et al.* Biomarkers associated with prostate cancer progression. *J. Cell. Biochem. (Suppl.)* **19**, 208–216 (1994).
- Grasso, A., Wen, D., Pretlow, T. & Kung, H. ErbB kinases and NDF signaling in prostate cancer cells. *Oncogene* **15**, 2705–2717 (1997).
- Holmes, W. E. *et al.* Identification of heregulin, a specific activator of p185erbB2. *Science* **256**, 1205–1210 (1992).
- Wen, D. *et al.* Neu differentiation factor: a transmembrane glycoprotein containing an EGF domain and an immunoglobulin homology unit. *Cell* **69**, 559–572 (1992).
- Dougall, W. C. *et al.* The neu-oncogene: signal transduction pathways, transformation mechanisms and evolving therapies. *Oncogene* **9**, 2109–2123 (1994).
- Levitzi, A. & Gazit, A. Tyrosine kinase inhibition: an approach to drug development. *Science* **267**, 1782–1788 (1995).
- Beerli, R. R., Wels, W. & Hynes, N. E. Intracellular expression of single chain antibodies reverts ErbB-2 transformation. *J. Biol. Chem.* **269**, 23931–23936 (1994).
- Carraway, K. L. R. & Cantley, L. C. A new acquaintance for erbB3 and erbB4: a role for receptor heterodimerization in growth signaling. *Cell* **78**, 5–8 (1994).
- Guy, P. M., Platko, J. V., Cantley, L. C., Cerione, R. A. & Carraway, K. L. R. Insect cell-expressed p180erbB3 possesses an impaired tyrosine kinase activity. *Proc. Natl Acad. Sci. USA* **91**, 8132–8136 (1994).
- Kishimoto, T., Taga, T. & Akira, S. Cytokine signal transduction. *Cell* **76**, 253–262 (1994).
- Taniguchi, T. Cytokine signaling through nonreceptor protein tyrosine kinases. *Science* **268**, 251–255 (1995).
- Chen-Kiang, S. Regulation of terminal differentiation of human B-cells by IL-6. *Curr. Topics Microbiol. Immunol.* **194**, 189–198 (1995).
- Fukada, T. *et al.* Two signals are necessary for cell proliferation induced by a cytokine receptor gp130: involvement of STAT3 in anti-apoptosis. *Immunity* **5**, 449–460 (1996).
- Siegal, C. B., Schwab, G., Nordan, R. P., FitzGerald, D. J. & Pastan, I. Expression of the interleukin 6 receptor and interleukin 6 in prostate carcinoma cells. *Cancer Res.* **50**, 7786–7788 (1990).
- Carpenter, G. & Cohen, S. Epidermal growth factor. *J. Biol. Chem.* **265**, 7709–7712 (1990).
- Rosette, C. & Karin, M. Ultraviolet light and osmotic stress: activation of the JNK cascade through multiple growth factor and cytokine receptors. *Science* **274**, 1194–1197 (1996).
- Daub, H., Weiss, F. U., Wallasch, C. & Ullrich, A. Role of transactivation of the EGF receptor in signalling by G-protein-coupled receptors. *Nature* **379**, 557–560 (1996).
- Yamauchi, T. *et al.* Tyrosine phosphorylation of the EGF receptor by the kinase Jak2 is induced by growth hormone. *Nature* **390**, 91–96 (1997).
- Rui, H., Kirken, R. A. & Farrar, W. L. Activation of receptor-associated tyrosine kinase JAK2 by prolactin. *J. Biol. Chem.* **269**, 5364–5368 (1994).

**Acknowledgements.** We thank N. E. Hynes for the scFv5R clone, J. Jacobberger and T. G. Pretlow for CWR22 cells, and A. W. Grasso and K. Everiss for reading the manuscript. This work was supported by NIH grants. Y.Q. is a recipient of an NCI research oncology training fellowship.

Correspondence and requests for materials should be addressed to H.J.K. (e-mail: hxx5@po.cwru.edu).

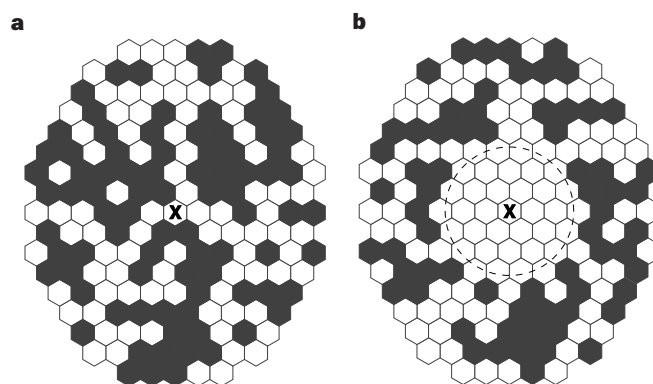
## Receptor clustering as a cellular mechanism to control sensitivity

Dennis Bray, Matthew D. Levin & Carl J. Morton-Firth

Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Chemotactic bacteria such as *Escherichia coli* can detect and respond to extremely low concentrations of attractants, concentrations of less than 5 nM in the case of aspartate<sup>1</sup>. They also sense gradients of attractants extending over five orders of magnitude in concentration (up to 1 mM aspartate)<sup>2,3</sup>. Here we consider the possibility that this combination of sensitivity and range of response depends on the clustering of chemotactic receptors on the surface of the bacterium<sup>4</sup>. We examine what will happen if ligand binding changes the activity of a receptor, propagating this change in activity to neighbouring receptors in a cluster<sup>5,6</sup>. Calculations based on these assumptions show that sensitivity to extracellular ligands increases with the extent of spread of activity through an array of receptors, but that the range of concentrations over which the array works is severely diminished. However, a combination of low threshold of response and wide dynamic range can be attained if the cell has both clusters and single receptors on its surface, particularly if the extent of activity spread can adapt to external conditions. A mechanism of this kind can account quantitatively for the sensitivity and response range of *E. coli* to aspartate.

The effects of activity spread may be illustrated by considering



**Figure 1** Activity spread in a cluster of receptors. Part of a hypothetical 1,000-receptor array is shown; individual receptor molecules are represented as white or grey hexagons, depending on their conformation. Receptors are supposed to switch randomly between a high-activity conformation (grey) and a low-activity conformation (white). **a**, One receptor in the centre of the array, marked with a cross, is occupied by a ligand molecule and consequently is in the low-activity conformation. **b**, In this array the ligand-bound, low-activity state of the central receptor has spread to 36 neighbouring unoccupied receptors. Because of this 'infectivity', the single ligand molecule produces a larger change in the array and the signal sent from the array is stronger. However, the range over which the array detects changes in the concentration of ligand is much smaller in **b** than in **a**.

a model array of 1,000 receptors with properties loosely based on the chemotactic receptors of coliform bacteria<sup>7,8</sup>. Each receptor is assumed to adopt either an active or an inactive conformation and to switch rapidly between the two on a submillisecond timescale. In its active conformation, a receptor signals to downstream elements of the pathway. The probability of occupying this conformation constitutes the 'activity' of the receptor. We will denote this activity by the symbol  $p$ , using  $p_0$  for the activity of a receptor not occupied by ligand and  $p_1$  for the activity of a receptor occupied by ligand. We will consider below the possibility that, because of cooperative interactions, unoccupied receptors can also adopt a probability  $p_1$ . Both  $p_0$  and  $p_1$  can change with adaptation, but we will assume that  $p_0$  is always greater than  $p_1$  so that ligand binding causes a decrease in activity, as it does in bacterial chemotaxis.

In the initial array of 'naïve' receptors, let  $p_0 = 0.5$ ; that is, we assume that roughly half of the receptors in the array are active at any time (Fig. 1a). The total activity of the array—the strength of the signal it sends to downstream elements of the signal pathway—will then be  $Np_0 = 500$  units (each 'unit' being the strength of signal produced by one receptor in its active conformation). If we now expose the array to molecules of ligand A, then some of these molecules ( $n_A$ ) will bind to individual receptors. The occupied receptors will reduce their probability of switching to an active state to  $p_1$  and the total activity of the array will become smaller ( $n_A p_1 + (N - n_A) p_0$ ). Thus, if 200 molecules of ligand A bind and if, with each binding, a receptor reduces its activity from 0.5 to zero ( $p_1 = 0$ ), then the total output will decrease from 500 to 400 ( $200 \times 0.0 + 800 \times 0.5$ ) units.

In this example, the change in total activity is given by the number of receptors that are newly bound to ligand A, multiplied by the change in activity per receptor. However, if each molecule A that binds to the array affects the switching probability of multiple contiguous receptors, then the same number of binding events will generate a larger change in activity (Fig. 1b). Specifically, if we assume that a single binding event 'infects'  $n_i$  receptors (changes their probability of being active from  $p_0$  to  $p_1$ ), then the change in output activity  $\Delta P$  will be (for low concentrations of ligand)  $\Delta P = n_A n_i \sigma$ , where  $n_A$  is the number of ligand molecules bound and  $\sigma (= p_0 - p_1)$  is the change in activity per receptor.

For an array of  $N$  receptors, each with a single binding site of dissociation constant  $K_d$  exposed to ligand A at a concentration  $C_A$ ,  $n_A/N = C_A/(C_A + K_d)$ . As the concentration of A becomes less, the change in activity will fall until it reaches a minimum detectable by experimental observation. Calling this minimum concentration  $C_A^{\min}$  and observing that  $n_A^{\min} = \Delta P^{\min}/n_i \sigma$ , then:

$$C_A^{\min} = \left( \frac{n_A^{\min}}{N - n_A^{\min}} \right) K_d = \left( \frac{\Delta P^{\min}/n_i \sigma}{N - \Delta P^{\min}/n_i \sigma} \right) K_d \quad (1)$$

From equation (1) we see that the minimum detectable concentration of ligand (the threshold concentration) decreases as the extent of activity spread grows larger. Indeed, if  $n_i$  is sufficiently large, then the entire array can switch state in response to the binding of a single ligand molecule. But if a single binding event fires the whole array, then the system will be unable to detect binding of subsequent ligand molecules. As this extreme example illustrates, the larger  $n_i$  is, the lower the range over which detection can occur.

To calculate the maximum detectable concentration of ligand, we must consider adaptation. The sensory systems of single cells, like those of entire organisms, adjust to ambient conditions by continually changing their baseline activity. This enables them to react to small changes even against a high background level of stimulation. The most effective form of adaptation (capable of extending over the widest range) occurs if the activity always returns to its original value—that is, it is exact<sup>9</sup>.

Applying exact adaptation, we see that, following exposure to ligand, the probabilities that receptors will be in the active conformation ( $p_0$  and  $p_1$ ) must rise so as to restore the original activity

(500 units in our example). We cannot say exactly how the probabilities will change but there are boundary conditions. Thus, if  $p_1$  remains zero,  $p_0 = 1$  and  $n_i = 1$ , then an output value of 500 units will be reached when just half of the receptors are occupied ( $500 \times 1 + 500 \times 0$ ). This will occur when the ambient concentration of ligand equals the  $K_d$  of the receptors, which is a very poor performance. It becomes even worse if  $p_0 < 1$  or  $n_i > 1$ .

A much wider range can be obtained by allowing both  $p_0$  and  $p_1$  to rise with increasing ligand concentration, subject to the constraint that the total activity of the array must be 500 units. That is,  $(1,000 - n_A) p_0 + n_A p_1 = 500$ . The widest possible range is reached by making  $p_0 = 1$ , and allowing  $p_1$  to rise as close as possible to 0.5; then the total activity of the array will be close to 500 even when all receptors are occupied ( $1,000 \times 0.5 = 500$ ). In practical terms, the upper limit in concentration of ligand will occur when the output activity is so close to 500 that the difference cannot be measured experimentally, that is, when addition of saturating levels of ligand A raises the activity by less than  $\Delta P^{\min}$ .

We must therefore calculate how many molecules of ligand A must bind to our array before the output activity comes within

## Box 1 Raindrops calculation

Consider an array of  $N$  membrane-bound receptors (represented by the grey area in the figure). Let some of these receptors bind to ligand A and let each binding event change the average activity of  $n_i$  receptors (white circles in the figure). What is the probability that any given receptor of the initial set of  $N$  remains unaffected after  $n_A$  ligand molecules have bound?

If one ligand molecule binds, then  $n_i$  receptors will change their activity. The probability that a given receptor is not affected by this event is given by the ratio:

$$\frac{N - n_i}{N} = \left( 1 - \frac{n_i}{N} \right)$$

A second ligand molecule has available to it  $N - 1$  sites (as one is already occupied). The number of receptor molecules changing their conformation in response to this second binding is again  $n_i$ , so that the probability that any given receptor is not affected by the second ligand is:

$$\frac{N - 1 - n_i}{N - 1} = \left( 1 - \frac{n_i}{N - 1} \right)$$

These two ligand binding events are independent. Therefore, the probability that a given receptor will not be affected by either binding event is the product of the two, namely:

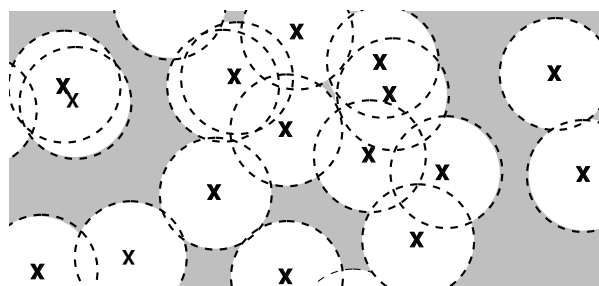
$$\left( 1 - \frac{n_i}{N} \right) \left( 1 - \frac{n_i}{N - 1} \right)$$

The same argument holds for each successive molecule of ligand A, so that the probability of being unaffected after  $n_A$  ligand molecules have bound is:

$$\left( 1 - \frac{n_i}{N} \right) \left( 1 - \frac{n_i}{N - 1} \right) \left( 1 - \frac{n_i}{N - 2} \right) \dots \left( 1 - \frac{n_i}{N - n_A + 1} \right)$$

The number of ligand molecules unaffected by the binding of  $n_A$  receptor molecules is given by the function R, where:

$$R = N \prod_{k=1}^{n_A} \left( 1 - \frac{n_i}{N - k + 1} \right)$$





$\Delta P^{\min}$  of maximum. Again we include the possibility of activity spread, but in this case we must allow for overlap between successive receptor bindings (the raindrop effect; see Box 1). The number of ligand molecules required can be deduced from the relationship  $\Delta P^{\min}/\sigma \leq R$ , where function  $R(N, n_i, n_A)$  is the number of receptors remaining unaffected after  $n_A$  molecules of ligand have bound (see Box 1) and  $\sigma (= p_0 - p_1)$  has its optimal value of 0.5. The value of  $R$  can be readily computed for increasing values of  $n_A$  until the above inequality is satisfied. This gives  $n_A^{\max}$ , from which we deduce the maximum detectable concentration of ligand  $C_A^{\max}$ :

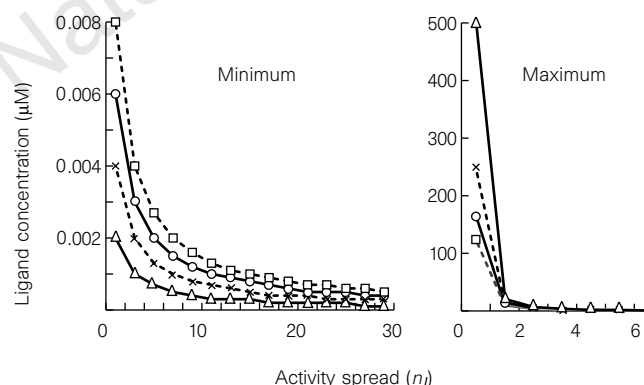
$$C_A^{\max} = \left( \frac{n_A^{\max}}{N - n_A^{\max}} \right) K_d \quad (2)$$

We thus have equations for the minimum and maximum ligand concentrations for a range of activity spreads (that is,  $n_i$  values) and for different levels of detectable activity  $\Delta P^{\min}$ . As seen in Fig. 2, the minimum concentration calculated from equation (1) falls in hyperbolic fashion with increasing  $n_i$ , an increase from  $n_i = 1$  to  $n_i = 2$  producing a twofold decrease in  $C_A^{\min}$ . However, the maximum concentration calculated from equation (2) falls very much more steeply. Here an increase from  $n_i = 1$  to  $n_i = 2$  produces a

**Table 1** Receptor activity level calculated from swimming performance

| Rotational bias | CheYp ( $\mu\text{M}$ ) | Output activity ( $P$ ) | Receptor activity |
|-----------------|-------------------------|-------------------------|-------------------|
| 0               | 15                      | 2,000                   | 1.0               |
| 0.75            | 2.8                     | 109                     | 0.054             |
| 0.85            | 2.5                     | 95                      | 0.048             |
| 0.95            | 2.0                     | 74                      | 0.037             |
| 1.0             | 0                       | 0                       | 0                 |

The rotational bias of an *E. coli* cell is the fraction of time the flagellum spends in a smooth-swimming (or anticlockwise motor rotation) mode. The bias of a wild-type cell without chemotactic stimulation is about 0.85 (refs 1, 19). Bias is controlled by the cytoplasmic concentration of CheYp and there is a cooperative relationship between the two, with a Hill coefficient of 5–6 (ref. 17). Here we use a relationship of the form  $\text{bias} = 1 - Y_p^{5.5} / (5.67(2.51)^{5.5} + Y_p^{5.5})$ , in which the steady-state concentration of CheYp ( $Y_p$ ) in a wild-type, unstimulated cell is 2.51  $\mu\text{M}$ . We take the lower limit of  $Y_p$  to be zero and the upper limit to be the concentration at which the bias falls below 0.001, namely 15  $\mu\text{M}$ . In a wild-type cell, the concentration of CheYp depends mainly on the rate of formation of CheYp by the receptor complex and the rate of breakdown of CheYp by CheZ. At steady state, these two balance, so that  $P(20 - Y_p) \propto ZY_p$ , where  $P$  is the output activity of the chemotactic receptors (as defined above),  $Z$  is the activity of CheZ, and the total concentration of CheY is taken to be 20  $\mu\text{M}$  (ref. 27). This leads to a relationship of the form  $P = \text{const}(Y_p / (20 - Y_p))$ , and we assign the constant a value of 2,000/3 so that  $P = 2,000$  at  $Y_p = 15 \mu\text{M}$  (bias of 0.001, receptors fully active) and  $P = 0$  at  $Y_p = 0.0 \mu\text{M}$  (bias of 1.0, receptors inactive). The average activity of individual receptors is then calculated as  $P/2,000$ .



**Figure 2** Changes in the response to threshold and range of ligand concentration with the extent of activity spread. Minimum and maximum ligand concentrations to which the model 1,000-receptor array can respond, with output values of 500 units, are plotted against activity spread ( $n_i$ ). In each graph, four different values of the minimum detectable change in output activity ( $\Delta P^{\min}$ ) are shown as follows: one unit, triangles connected by complete lines; two units, crosses connected by dashed lines; three units, circles connected by complete lines; and four units, squares connected by dashed lines. Note the differences in scale of the two graphs.

greater than 20-fold drop in the maximum (saturating) concentration of ligand A. Thus, activity spread lowers the threshold but severely restricts the dynamic range over which the receptor array can operate.

Let us now apply these ideas to bacterial chemotaxis and in particular to the system by which *E. coli* detect and swim towards distant sources of aspartate. Each cell has ~2,000 dimeric aspartate receptors on its surface, many of them clustered into a patch at the polar regions of the cell<sup>4,10,11</sup>. The receptors are part of a complex of proteins, including the autophosphorylating kinase CheA, which transfers phosphoryl groups onto CheY<sup>12</sup>. Phosphorylated CheY (CheYp) molecules diffuse to the flagellar motors, causing them to rotate in a clockwise direction and thereby producing a random change in direction of the bacterium<sup>13</sup>. The rate at which phosphate groups are generated by the receptor complex falls when aspartate is bound<sup>14</sup>, causing the motors to turn anticlockwise and the cell to persist in its current direction. In effect, the flux of phosphoryl groups from the receptor complex constitutes its output activity, equivalent to the array activity  $P$  mentioned above.

Much quantitative information is available relating to the biochemical steps of this chemotactic signal pathway<sup>7</sup>. In particular, the rates of phosphotransfer from the receptor complex to CheY and the rate of loss of phosphate groups from CheYp (mainly through CheZ) are known<sup>15,16</sup>. We also have estimates of the relation between the steady-state concentration of CheYp and the flagellar rotation bias (the time spent in a smooth swimming mode)<sup>17</sup>. These experimentally measured values allow us to estimate the (hypothetical) switching frequency of aspartate receptors corresponding to different swimming biases (Table 1).

The values in Table 1 give an estimate of the change in output activity,  $\Delta P^{\min}$ , needed to produce a bias change of 0.1, which is the smallest detectable value for tethered bacteria. With  $\Delta P^{\min}$  we can calculate the smallest number of receptors that must change their average activity from 0.048 (the resting value) to zero, and hence the smallest concentration of aspartate that can be detected. In the absence of activity spread, this value is about 0.22  $\mu\text{M}$  or 220 nM, a value close to that given by computer models of the chemotaxis pathway on the basis of experimentally determined rates and concentrations in the chemotaxis pathway<sup>18,19</sup>.

However, the actual threshold of the chemotactic system is much lower than indicated by this figure. Classic experiments in which *E. coli* cells tethered to a coverslip were exposed to small quantities of chemoattractant delivered iontophoretically indicated that a change in receptor occupancy of as little as 1/600 could produce a detectable change in swimming behaviour<sup>1</sup>. With a  $K_d$  of 1  $\mu\text{M}$ <sup>20,21</sup>, this change corresponds to a minimum detectable concentration of about 2 nM aspartate. Estimates obtained using caged aspartate<sup>22</sup> also show a response threshold corresponding to about 0.002 occupancy, or ~2 nM (R. Jasuja and S. Khan, manuscript in preparation). Another difficulty with current models of the chemotaxis pathway concerns the response to substances such as ribose that are detected by receptors present in small numbers. The ribose receptor, Trg, accounts for only ~1% of the total chemotactic receptors of the cell and yet operates through the same phosphorylation cascade as the aspartate receptors<sup>23</sup>. On a simple 'head-count' mechanism, signals from Trg receptors would simply not be heard.

Both discrepancies between theory and experiment can be resolved if activities can spread from one receptor to its neighbours in the cluster. Using the equations already given, we can readily achieve the required threshold to both aspartate or ribose, given sufficient activity spread. But, as before, we must also consider the range of detection, which become less as the infectivity grows. Information on the upper limits of chemotactic detection is difficult to find: the best estimates are probably those obtained using a capillary accumulation assay<sup>2</sup>. For the non-hydrolysable analogue of aspartate,  $\alpha$ -methylaspartate, concentrations in the range from 0.3  $\mu\text{M}$  to 0.1 M produced chemotactic responses. As the binding

**Table 2 Effect of receptor clustering on chemotactic threshold and range**

| No. of receptors free | No. of receptors clustered | Activity spread ( $n$ ) | Minimum conc. ( $\mu\text{M}$ ) | Maximum conc. ( $\mu\text{M}$ ) |
|-----------------------|----------------------------|-------------------------|---------------------------------|---------------------------------|
| 2,000                 | 0                          | 1                       | 0.22                            | 730                             |
| 1,000                 | 1,000                      | 1                       | 0.22                            | 730                             |
| 1,000                 | 1,000                      | 10                      | 0.038*                          | 360†                            |
| 1,000                 | 1,000                      | 100                     | 0.004*                          | 360†                            |
| 0                     | 2,000                      | 1                       | 0.22                            | 730                             |
| 0                     | 2,000                      | 10                      | 0.019                           | 4.2                             |
| 0                     | 2,000                      | 100                     | 0.002                           | 0.33                            |

Experimental data on the chemotaxis of *E. coli* were used with equations (1) and (2) to estimate the minimum and maximum detectable concentrations of aspartate. The minimum detectable concentration of aspartate,  $C^{\text{min}}$  (also referred to as the threshold), was calculated from equation (1) using a  $K_d$  of 1  $\mu\text{M}$  for a non-adapted aspartate receptor, a value of  $\Delta P^{\text{min}} = 17.5$  obtained from the data in Table 1, and a change in activity associated with individual receptors of  $\sigma = 1 - 0.048 = 0.95$ . The maximum detectable concentration of aspartate,  $C^{\text{max}}$  (also referred to as the saturating concentration, or chemotactic range), was calculated using a  $K_d$  for the binding of aspartate to the fully methylated (fully adapted) chemotactic receptor of 7  $\mu\text{M}$  (refs 24, 25) and values of  $R$  calculated as described in Box 1.

\* Contribution from free receptors assumed to be negligible.

† Contribution from clustered receptors assumed to be negligible.

of methylaspartate to the aspartate receptor is relatively weak ( $K_d$  close to 0.1 mM), we estimate that aspartate itself (which has a  $K_d$  of 1  $\mu\text{M}$ ) should give the same change in receptor occupancy in the range of 3 nM to 1 mM.

In Table 2 we give the expected minimum and maximum concentrations of aspartate detected by a set of 2,000 aspartate receptors, making different assumptions about their degree of clustering and activity spread. These values also include an increase in the  $K_d$  for aspartate with adaptation<sup>24,25</sup>. We see that if half of the receptors are freely diffusing and the other half are clustered and show an infectivity of 100, aspartate will be detected in the concentration range 4 nM to 360  $\mu\text{M}$ . However, the best overall performance is achieved if receptors are totally aggregated at low concentrations of aspartate and totally non-aggregated at high concentrations. In this case, the cell can achieve a range of detection at concentrations from 2 nM to 730  $\mu\text{M}$ , a range of greater than five orders of magnitude and similar to the best estimates of the actual performance. Changes of this kind might be produced, for example, by the methylating enzyme CheR which can link adjacent receptors and whose binding is reduced by methylation<sup>11,26</sup>.

The model we have proposed requires that when chemotactic receptors cluster on the cell surface, the activity of one receptor influences that of its neighbours. Furthermore, to obtain a combination of sensitivity and range of response the cell must be able to control either the aggregation itself or the degree to which the receptor activity can propagate. We predict, for example, that aggregation (or activity spread) will be lowest when the cell has adapted to high concentrations of ambient attractant. We also expect that chemoattractants such as ribose that operate through minor receptors will saturate at very low concentrations (as they can be sensed only through activity spread). There should be conditions in which the degree of aggregation, and hence the sensitivity to chemoattractants, is impaired (this might occur, for example, in mutants lacking CheR and CheB). Finally, given the simplicity of this mechanism, we anticipate that it will be widely used by cells other than bacteria and for purposes other than chemotaxis. □

Received 12 January; accepted 26 February 1998.

- Segall, J. E., Block, S. M. & Berg, H. C. Temporal comparisons in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **83**, 8987–8991 (1986).
- Mesibov, R., Ordal, G. W. & Adler, J. The range of attractant concentrations for bacterial chemotaxis and the threshold and size of response over this range. Weber law and related phenomena. *J. Gen. Physiol.* **62**, 203–223 (1973).
- Berg, H. C. & Tedesco, P. M. Transient response to chemotactic stimuli in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **72**, 3235–3239 (1975).
- Maddock, J. R. & Shapiro, L. Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259**, 1717–1723 (1993).
- Gardina, P. J. & Manson, M. D. Attractant signaling by an aspartate chemoreceptor dimer with a single cytoplasmic domain. *Science* **274**, 425–426 (1996).
- Parkinson, J. S. & Blair, D. F. Does *E. coli* have a nose? *Science* **259**, 1701–1702 (1993).
- Stock, J. B. & Surette, M. G. In *Escherichia and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 1103–1129 (Am. Soc. Microbiol., Washington DC, 1996).
- Eisenbach, M. Control of bacterial chemotaxis. *Mol. Microbiol.* **4**, 161–167 (1996).

- Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
- Gegner, J. A., Graham, D. R., Roth, A. F. & Dahlquist, F. W. Assembly of an MPC receptor, CheW, and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. *Cell* **70**, 975–982 (1992).
- Li, J. Y., Li, G. Y. & Weis, R. M. The serine chemoreceptor from *Escherichia coli* is methylated through an interdimer process. *Biochemistry* **36**, 11851–11857 (1997).
- Schuster, S. C., Swanson, R. V., Alex, L. A., Bourret, R. B. & Simon, M. I. Assembly and function of a quaternary signal transduction complex monitored by surface plasmon resonance. *Nature* **365**, 343–346 (1993).
- Barak, R. & Eisenbach, M. Correlation between phosphorylation of the chemotaxis protein CheY and its activity at the flagellar motor. *Biochemistry* **31**, 1821–1826 (1992).
- Ninfa, E. G., Stock, A., Mowbray, S. & Stock, J. Reconstitution of the bacterial chemotaxis signal transduction system from purified components. *J. Biol. Chem.* **266**, 9764–9770 (1991).
- Stewart, R. C. Activating and inhibitory mutations in the regulatory domain of CheB, the methyltransferase in bacterial chemotaxis. *J. Biol. Chem.* **266**, 1921–1930 (1993).
- Lukat, G. S., McCleary, W. R., Stock, A. M. & Stock, J. B. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *Proc. Natl Acad. Sci. USA* **89**, 718–722 (1992).
- Kuo, S. C. & Koshland, D. E. Jr Multiple kinetic states for the flagellar motor switch. *J. Bacteriol.* **171**, 6279–6287 (1989).
- Bray, D., Bourret, R. B. & Simon, M. I. Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol. Biol. Cell* **4**, 469–482 (1993).
- Levin, M. D., Morton-Firth, C. J., Abouhamad, W. N., Bourret, R. B. & Bray, D. Origins of individual swimming behavior in bacteria. *Biophys. J.* **74**, 175–181 (1998).
- Biemann, H.-P. & Koshland, D. E. Aspartate receptors of *Escherichia coli* and *Salmonella typhimurium* bind ligand with negative and half-of-sites cooperativity. *Biochemistry* **33**, 629–634 (1994).
- Aksent, R. R., Howlett, B. J. & Koshland, D. E. Jr Soluble and membrane-bound aspartate-binding activities in *Salmonella typhimurium*. *J. Bacteriol.* **123**, 1000–1005 (1975).
- Khan, S. et al. Excitatory signaling in bacteria probed by caged chemoeffectors. *Biophys. J.* **65**, 2368–2382 (1993).
- Hazelbauer, G. L., Engstrom, P. & Harayama, S. Methyl-accepting chemotaxis protein III and transducer gene *trg*. *J. Bacteriol.* **145**, 43–49 (1981).
- Borkovich, K. A., Alex, L. A. & Simon, M. I. Attenuation of sensory receptor signaling by covalent modification. *Proc. Natl Acad. Sci. USA* **89**, 6756–6760 (1992).
- Yonekawa, H. & Hayashi, H. Desensitization by covalent modification of the chemoreceptor of *Escherichia coli*. *FEBS Lett.* **198**, 21–24 (1986).
- Wu, J., Li, J., Li, G., Long, D. G. & Weis, R. M. The receptor binding site for the methyltransferase of bacterial chemotaxis is distinct from the sites of methylation. *Biochemistry* **35**, 4984–4993 (1996).
- Stock, A. M., Koshland, D. E. Jr & Stock, J. B. Homologies between the *Salmonella typhimurium* CheY protein and proteins involved in the regulation of chemotaxis, membrane protein synthesis, and sporulation. *Proc. Natl Acad. Sci. USA* **82**, 7989–7993 (1985).

**Acknowledgements.** We thank M. Keeling for help with mathematics, S. Khan for permission to cite unpublished data, and R. Bourret, E. Siggia and T. Lamb for criticisms of the manuscript.

Correspondence and requests for materials should be addressed to D.B. (e-mail: d.bray@zoo.cam.ac.uk).

## Chromatin remodelling by the glucocorticoid receptor requires the BRG1 complex

Christy J. Fryer & Trevor K. Archer

Departments of Obstetrics & Gynaecology, Biochemistry and Oncology, University of Western Ontario, London Regional Cancer Centre, 790 Commissioners Rd East, London, Ontario N6A 4L6, Canada

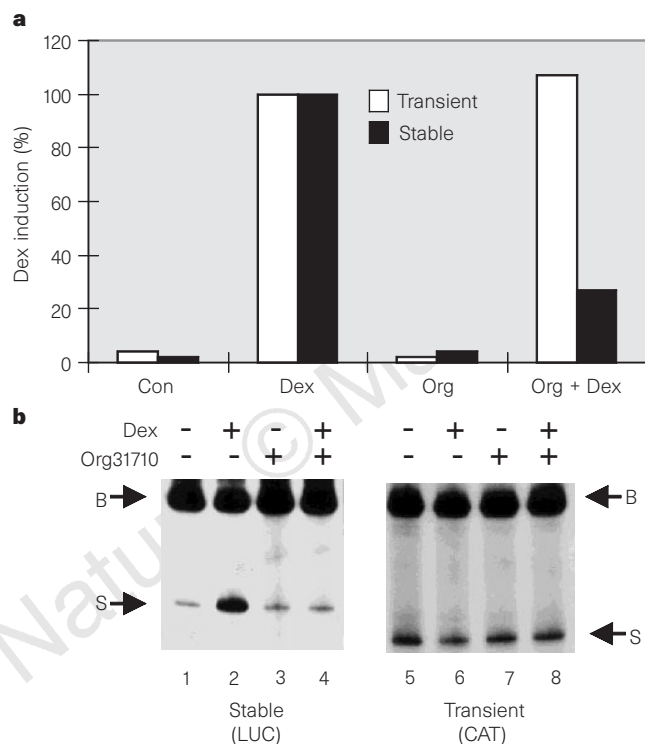
The assembly of transcriptional regulatory DNA sequences into chromatin plays a fundamental role in modulating gene expression<sup>1,2</sup>. The promoter of the mouse mammary-tumour virus (MMTV) is packaged into a regular array of nucleosomes when it becomes stably integrated into mammalian chromosomes, and has been used to investigate the relationship between chromatin architecture and transcriptional activation by the hormone-bound glucocorticoid and progesterone receptors<sup>3,4</sup>. In mammalian cells that express both of these receptors, the progesterone receptor activates transcription from transiently transfected MMTV DNA<sup>5,6</sup> but not from organized chromatin templates<sup>7</sup>. Moreover, the activated progesterone receptor inhibits the chromatin remodelling and consequent transcriptional stimulation that is mediated by the glucocorticoid receptor. Here we investigate the mechanism of this inhibition by characterizing the interaction of the glucocorticoid receptor with transcriptional co-activator and chromatin remodelling protein complexes<sup>2,8</sup>. We show that when this receptor is prevented from interacting with the hBRG1/BAF chromatin remodelling complex, it can activate transcription from transiently transfected

**DNA but not from organized chromatin templates. Our results indicate that it may be possible to separate the transcriptional activation and chromatin remodelling activities of proteins that interact with hormone receptors.**

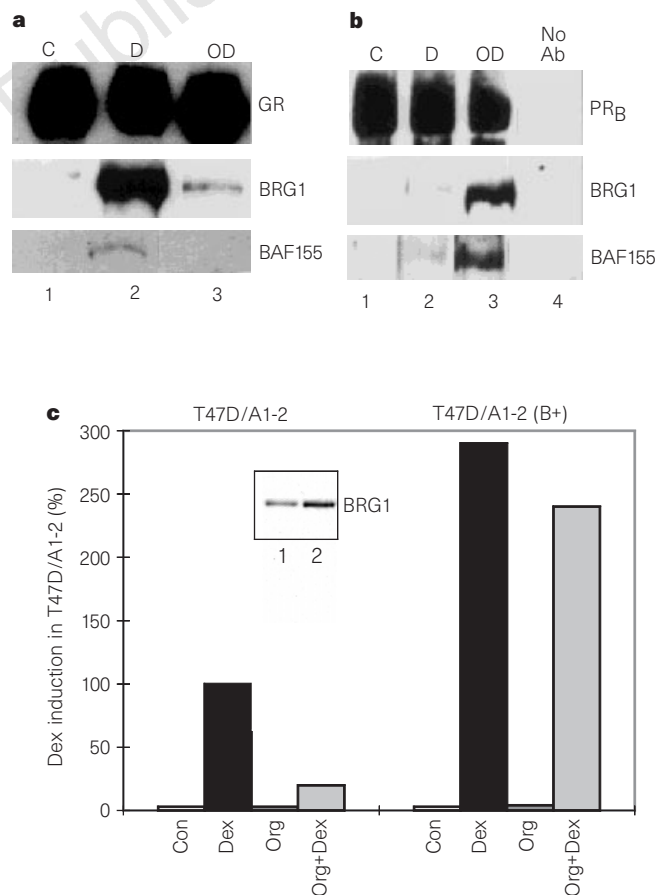
To investigate how the progesterone receptor (PR) inhibits the activity of the glucocorticoid receptor (GR), we tested whether it could also inhibit GR activity when bound to an antiprogesterin. We used the antiprogesterin Org31710, which is highly specific for the PR rather than the GR<sup>9</sup>, at a 10-fold lower concentration than dexamethasone (Dex) (Dex,  $10^{-7}$  M; Org31710,  $10^{-8}$  M). Org31710 inhibited the GR-mediated activation of the stable template but not of the transiently transfected template (Fig. 1a). Consistent with this, antiprogesterin blocked the glucocorticoid-induced chromatin remodelling, as measured by hypersensitivity to digestion by restriction enzymes (Fig. 1b, lanes 2 and 4). As expected from the continued transcription from the transient template in the presence of antiprogesterin, transiently transfected MMTV DNA is equally accessible to restriction enzymes in the presence and absence of hormone (Fig. 1b). The differences between the transient and chromatin templates arise because transient templates are not assembled into a phased array of nucleosomes and so are constitutively

hypersensitive and are bound by *trans*-acting factors<sup>5,6</sup>. As a control for the specificity of Org31710 in cells lacking the PR, we found that the antiprogesterin had no effect on GR-mediated chromatin remodelling or transcriptional activation of the MMTV promoter (C.F. and T.K.A., data not shown). These results show that the PR does not need to be transcriptionally competent to inhibit the GR and indicate that inhibition occurs at the initial chromatin-remodelling step.

We next investigated how the PR prevents the GR from remodelling organized chromatin. Given that the promoter must be remodelled for transcriptional activation<sup>7,10</sup> and that the SWI-SNF/hBRG1 complex is needed for GR *trans*-activation<sup>11-13</sup>, we tested for interaction between the GR and the hBRG1 complex. Co-immunoprecipitation experiments showed that glucocorticoids promote a hormone-dependent association of the GR with the hBRG1 complex (Fig. 2a, lanes 1 and 2). Consistent with Org31710's ability to block chromatin remodelling, hormone-dependent GR/hBRG1 association is inhibited by this antiprogesterin (Fig. 2a, lanes 2 and 3). In mammalian cells, BRG1 is complexed with other proteins, BAFs for example<sup>14</sup>; we found that Org31710 inhibits the association of GR with BAF 155 (Fig. 2a, lanes 2 and 3) and with BAF 60a



**Figure 1** Antiprogesterin inhibits GR-mediated chromatin remodelling and activation of stable but not transiently introduced MMTV templates. **a**, T47D/A1-2 cells were transiently transfected with  $10 \mu\text{g}$  of pHHCAT (MMTV-CAT) and CAT and luciferase were assayed to quantify transcription from the stable (black bars) and transient (white bars) MMTV templates. Cells were untreated (Con), treated with Org31710 ( $10^{-8}$  M) for 42 h (Org), treated with dexamethasone (Dex) ( $10^{-7}$  M) for 24 h (Dex) or pretreated with Org31710 for 18 h before Dex addition for 24 h (Org + Dex). Data are averaged from two experiments done in duplicate. The level of transcriptional induction by Dex was set to 100% and the other values are given as a percentage of Dex induction. **b**, T47D/A1-2 cells were transiently transfected with pHHCAT as in **a**, and were untreated, treated with Dex for 6 h, treated with Org31710 for 24 h, or pretreated with Org31710 for 18 h before dexamethasone addition for 6 h. Nuclei were isolated and digested with *SstI* (S). After purification, DNA was digested with *BamHI* (B) to provide an internal standard and as a control for the degree of enzyme accessibility *in vivo*. DNA was analysed as described<sup>6</sup> with primers specific for the stable (LUC) and transient (CAT) templates<sup>6</sup>. Prolonged incubation (24 h) with antiprogesterins is not necessary but does aid inhibition of GR by PR.



**Figure 2** Antiprogesterin inhibition of glucocorticoid-dependent association of GR with hBRG1 and BAF 155 is reduced by overexpression of hBRG1. **a**, Immunoprecipitation with an anti-GR antibody and immunoblotting of GR, hBRG1 and BAF 155 from T47D/A1-2 cells. Conditions were as for Fig. 1b, control (lane 1), Dex for 6 h (lane 2), or Org31710 for 18 h plus Dex for 6 h (lane 3). **b**, Immunoprecipitation with an anti-PR<sub>B</sub> antibody and immunoblotting of PR<sub>B</sub>, hBRG1 and BAF 155 from T47D/A1-2 cells. Lane assignments and treatments are as in **a**. Immunoprecipitation with no antibody (lane 4) as a negative control shows that the interaction is specific. **c**, Luciferase was assayed in T47D/A1-2 or T47D/A1-2(B+) cells to quantify transcription from stable MMTV templates. Data were derived as for Fig. 1a. Insert: Immunoblot of  $80 \mu\text{g}$  total cellular extracts from T47D/A1-2 (lane 1) and T47D/A1-2(B+) (lane 2) cells for hBRG1.

and BAF 170 (C.F. and T.K.A., unpublished results). To determine how the PR disrupts the hBRG1/GR complex, we did reciprocal experiments using a hPR<sub>B</sub> antibody. A PR/hBRG1 complex was evident in the presence of the antiprogesterin, suggesting that antiprogesterin/PR competes with dexamethasone/GR for the hBRG1 complex (Fig. 2b, lanes 2 and 3). Likewise, antiprogesterin/PR competes with the dexamethasone/GR for BAF 155 to form a PR/BAF 155 complex (Fig. 2b, lanes 2 and 3). These results are consistent with a model in which the GR/hBRG1 complex is required for the hormone-dependent remodelling of chromatin and subsequent *trans*-activation of the promoter<sup>11–13</sup>. The inability of antiprogesterin/PR to inhibit GR activation of the transient template was expected because transient templates do not require remodelling because of their constitutively open conformation<sup>5,6</sup>.

Steroid receptors associate with a variety of other proteins known as co-activators or co-integrators<sup>8,15</sup>. To determine the role of these proteins in the activation process and in the selective inhibitory effect of the PR upon organized chromatin compared with transiently transfected DNA, we investigated the interactions of some of these proteins with the GR. As reported previously, there is a hormone-dependent association of SRC-1/NcoA1 and GRIP-1/TIF-2/NcoA2 (refs 16–18) with the GR (Fig. 3a, lanes 1 and 2). In contrast to hBRG1, however, the antiprogesterin, which prevents the GR activation of organized chromatin, failed to disrupt the GR/co-activator complexes (Fig. 3a; compare lanes 2 and 3). However, the GR/co-activator complex, in the absence of hBRG1, can stimulate a transiently transfected MMTV reporter (Fig. 1a). This is consistent with the proposed role of these proteins in activating transcrip-

tion<sup>16–18</sup> and suggests that GR complexes that stimulate transcription may be separable from those that remodel chromatin.

The co-integrator protein p300/CBP is also important in the activation of transcription by steroid receptors<sup>8,16</sup>. Co-immunoprecipitation with an anti-GR antibody and immunoblotting with an anti-p300 antibody show that its association profile is identical to that seen with SRC-1/NcoA1 and GRIP-1/TIF-2/NcoA2 (Fig. 3b). The GR and p300/CBP undergo hormone-dependent association which is not perturbed by the antiprogesterin/PR complex, consistent with p300/CBP having a role in *trans*-activation, rather than in chromatin remodelling of the MMTV promoter (Fig. 3b, lanes 2–4). This is interesting as p300/CBP and associated proteins have histone acetyltransferase activity and so might be expected to play a role in chromatin remodelling<sup>19,20</sup>; however, MMTV and several other steroid-responsive genes are unusual in that they are not activated when histones are hyperacetylated<sup>21,22</sup>. Thus a GR/p300/CBP complex might not remodel this promoter, although p300/CBP and the co-activators may participate with a hBRG1/BAF complex in remodelling chromatin.

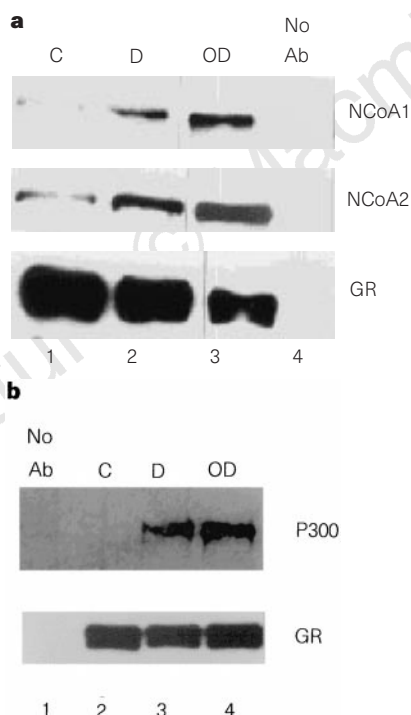
If, as indicated by our results, inhibition by PR of GR action is a result of competition for limiting quantities of hBRG1, it could be overcome by increasing the amount of hBRG1 in the cell. We therefore generated stable derivatives of T47D/A1-2 cells containing increased levels of hBRG1. Western blotting revealed that T47D/A1-2(B+) cells express 2.5-fold more hBRG1 than parental cells (Fig. 2c, insert), so we investigated the effect of this increased hBRG1 expression on the GR activation of chromatin and on the inhibition by the PR of this activation. As expected, increasing hBRG1 increased glucocorticoid induction 2.5–3-fold in response to dexamethasone (Fig. 2c; compare Con and Dex). Under conditions in which the PR inhibits GR activation by 80% in parental T47D/A1-2 cells, in T47D/A1-2(B+) cells containing more hBRG1 the PR no longer prevents GR induction (Fig. 2c; compare Dex and Org + Dex).

By analysing the protein partners for the activated GR, one can provide an explanation for the divergent activation of organized chromatin and transiently transfected MMTV templates *in vivo*<sup>5–7</sup>. We have shown that an activated GR complex includes at least p300/CBP, SRC-1/NcoA1, GRIP-1/TIF-2/NcoA2, Baf 155 and hBRG1, but it is not yet clear whether these interact directly with the GR or if one of these proteins acts as a scaffold for the complex. p300/CBP is a candidate scaffold protein: although no hBRG1-binding region has been identified, the GR, SRC-1/NcoA1 and GRIP-1/TIF-2/NcoA2 are known to bind to distinct regions of the molecule<sup>23–26</sup>. We have focused here on hBRG1, but there are other chromatin remodelling complexes and coactivators that may contribute to the transcriptional activation of organized chromatin templates<sup>27</sup>.

Our results demonstrate that hormone-dependent activation of the MMTV promoter by the GR requires the hBRG1/BAF complex. Under conditions that cause this complex to dissociate, *trans*-activation from an organized chromatin template is blocked. This inhibition can be reversed by overexpression of hBRG1, which enhances activation by the GR. Transactivation of transiently introduced templates is unaffected by dissociation of hBRG1 from GR, probably because these templates do not adopt a phased-nucleosome arrangement<sup>5,6</sup>. Transactivation of these templates depends on the association of the GR with co-activators such as SRC-1/NcoA1, GRIP-1/TIF-2/NcoA2 and p300/CBP. As a GR without hBRG1/BAFs can activate a transient but not an organized chromatin MMTV template, MMTV activation may depend on remodelling of the promoter as the first step in the activation of transcription *in vivo*. □

## Methods

**Cell lines.** T47D/A1-2 cells were derived from T47D breast cancer cells by stable transfection with plasmids pGRneo and pHHLuc as described<sup>7</sup>. T47D/A1-2 cells were grown at 37°C with 5% CO<sub>2</sub> in modified Eagle's medium



**Figure 3** Glucocorticoid-dependent association of GR with SRC-1/NcoA1, GRIP-1/TIF-2/NcoA2 and CBP/p300 is not inhibited by antiprogesterin. **a**, Immunoprecipitation with anti-GR antibody and immunoblotting with antibodies against NcoA1, NcoA2 and GR. T47D/A1-2 cells were treated as for Fig. 1b; control (C, lane 1), Dex ( $10^{-7}$  M) (D, lane 2) or Org31710 ( $10^{-8}$  M) plus Dex (OD, lane 3). Immunoprecipitation without antibody (lane 4) as a negative control shows that the interaction is specific. Western blot analysis of extracts from control and Dex-treated cells indicated that the overall amounts of these proteins were unchanged by hormone treatment (C.F. and T.K.A., unpublished results). **b**, Immunoprecipitation with anti-GR antibody and immunoblotting with antibodies against p300 (Santa Cruz) and GR. Treatments were as in **a**; no antibody, lane 1; control, lane 2; Dex, lane 3; and Org31710 plus Dex, lane 4.



containing 10% fetal bovine serum. T47D/A1-2 cells were transfected with 5  $\mu$ g pBJ5/BRG1a and 1  $\mu$ g of pGK puro and selected with 1  $\mu$ g ml<sup>-1</sup> puromycin<sup>28</sup>. Individual clones were selected and the presence of the hBRG1 expression plasmid confirmed by PCR. This cell line will be referred to as T47D/A1-2(B+). **CAT and luciferase assays.** For CAT and luciferase assays, subconfluent cells were plated one day before transfection with 10  $\mu$ g pHHCAT by lipofectamine (Gibco-BRL) according to the manufacturer's instructions. Fresh modified Eagle's medium with 10% fetal bovine serum was added 24 h before hormone treatment. Cells were collected, lysed and assayed for CAT and luciferase activity. CAT activity was determined by a kinetic assay<sup>7</sup>, and values were normalized to total protein. Luciferase was assayed according to the manufacturer's (Promega) protocol and normalized to total protein.

**Immunoprecipitation and immunoblotting.** For immunoprecipitation, cells were treated, washed with PBS and pelleted. The cells were lysed in buffer X (100 mM Tris-HCl, pH 8.5, 250 mM NaCl, 1% (v/v) NP-40, 1 mM EDTA, 1  $\mu$ g ml<sup>-1</sup> aprotinin 2 mg ml<sup>-1</sup> BSA). The supernatants were precleared once with 3% (w/v) protein A-Sepharose (Pharmacia) in buffer X. An anti-GR antibody, BUGR2 (ref. 29), or anti-PR<sub>B</sub> antibody, B-30 (ref. 30), was incubated with the supernatant for 1 h at 4°C, then with protein A-Sepharose for 1 h. Bound antibody/antigen complexes were washed 3 times in HEGNDT buffer (10 mM HEPES, pH 8.0, 1 mM EDTA, 10% glycerol, 50 mM NaCl, 2 mM DTT) containing 0.1% Triton X-100, and once in HEGNDT buffer. Washed immune complexes were resuspended in 2 × SDS-PAGE loading buffer and released by boiling for 5 min. For immunoblot analysis, proteins were solubilized in loading buffer, subjected to SDS-PAGE (7.5% gels) and transferred to nitrocellulose.

Received 17 November 1997; accepted 9 February 1998.

1. Grunstein, M. Histone function in transcription. *Annu. Rev. Cell. Biol.* **6**, 643–678 (1990).
2. Kingston, R. E., Bunker, C. A. & Imbalzano, A. N. Repression and activation by multiprotein complexes that alter chromatin structure. *Genes Dev.* **10**, 905–920 (1996).
3. Richard-Foy, H. & Hager, G. L. Sequence-specific positioning of nucleosomes over the steroid-inducible MMTV promoter. *EMBO J.* **6**, 2321–2328 (1987).
4. Archer, T. K. *et al.* Steroid hormone receptor status defines the MMTV promoter chromatin structure *in vivo*. *J. Ster. Biochem. Mol. Biol.* **53**, 421–429 (1995).
5. Archer, T. K., Lefebvre, P., Wolford, R. G. & Hager, G. L. Transcription factor loading on the MMTV promoter: A bimodal mechanism for promoter activation. *Science* **255**, 1573–1576 (1992).
6. Lee, H.-L. & Archer, T. K. Nucleosome-mediated disruption of transcription factor-chromatin initiation complexes at the mouse mammary tumor virus long terminal repeat *in vivo*. *Mol. Cell. Biol.* **14**, 32–41 (1994).
7. Archer, T. K., Zaniewski, E., Moyer, M. L. & Nordeen, S. K. The differential capacity of glucocorticoids and progestins to alter chromatin structure and induce gene expression in human breast cancer cells. *Mol. Endocr.* **8**, 1154–1162 (1994).
8. Glass, C. K., Rose, D. W. & Rosenfeld, M. G. Nuclear receptor coactivators. *Curr. Opin. Cell Biol.* **9**, 222–232 (1997).
9. Mizutani, T., Bhakta, A., Kloosterboer, H. J. & Moudgil, V. K. Novel antiprogesterone ORG 31806 and 31710: Interaction with mammalian progesterone receptor and DNA binding of antisteroid receptor complexes. *J. Ster. Biochem. Mol. Biol.* **42**, 695–704 (1992).
10. Truss, M. *et al.* Hormone induces binding of receptors and transcription factors to a rearranged nucleosome on the MMTV promoter *in vivo*. *EMBO J.* **14**, 1737–1751 (1995).
11. Yoshinaga, S. K., Peterson, C. L., Herskowitz, I. & Yamamoto, K. R. Roles of SWI1, SWI2, and SWI3 proteins for transcriptional enhancement by steroid receptors. *Science* **258**, 1598–1604 (1992).
12. Muchardt, C. & Yaniv, M. A human homologue of *Saccharomyces cerevisiae* SNF2/SWI2 and *Drosophila* brm genes potentiates transcriptional activation by the glucocorticoid receptor. *EMBO J.* **12**, 4279–4290 (1993).
13. Östlund Farrang, A.-K., Blomquist, P., Kwon, H. & Wrangé, Ö. Glucocorticoid receptor–glucocorticoid response element binding stimulates nucleosome disruption by the SWI/SNF complex. *Mol. Cell. Biol.* **17**, 895–905 (1997).
14. Wang, W. *et al.* Diversity and specialization of mammalian SWI/SNF complexes. *Genes Dev.* **10**, 2117–2130 (1996).
15. Horwitz, K. B. *et al.* Nuclear receptor coactivators and corepressors. *Mol. Endocr.* **10**, 1167–1177 (1996).
16. Kamei, Y. *et al.* A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* **85**, 403–414 (1996).
17. Oñate, S. A., Tsai, S. Y., Tsai, M.-J. & O'Malley, B. W. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* **270**, 1354–1357 (1995).
18. Voegel, J. J. *et al.* TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors. *EMBO J.* **15**, 3667–3675 (1996).
19. Ogryzko, V. V. *et al.* The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–959 (1996).
20. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
21. Bresnick, E. H. *et al.* Glucocorticoid receptor-dependent disruption of a specific nucleosome on the mouse mammary tumor virus promoter is prevented by sodium butyrate. *Proc. Natl Acad. Sci. USA* **87**, 3977–3981 (1990).
22. McKnight, G. S., Hager, L. & Palmiter, R. D. Butyrate and related inhibitors of histone deacetylation block the induction of egg white genes by steroid hormones. *Cell* **22**, 469–477 (1980).
23. Torchia, J. *et al.* The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature* **387**, 677–684 (1997).
24. Smith, C. L., Oñate, S. A., Tsai, M.-J. & O'Malley, B. W. REB binding protein acts synergistically with steroid receptor coactivator-1 to enhance steroid receptor-dependent transcription. *Proc. Natl Acad. Sci. USA* **93**, 8884–8888 (1996).
25. Chakravarti, D. *et al.* Role of CBP/P300 in nuclear receptor signalling. *Nature* **383**, 99–103 (1996).

26. Hong, H. *et al.* GRIP1, a novel mouse protein that serves as a transcriptional coactivator in yeast for the hormone binding domains of steroid receptors. *Proc. Natl Acad. Sci. USA* **93**, 4948–4952 (1996).
27. Wu, C. Chromatin remodeling and the control of gene expression. *J. Biol. Chem.* **272**, 28171–28174 (1997).
28. Dunaief, J. L. *et al.* The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. *Cell* **79**, 119–130 (1994).
29. Gametchu, B. & Harrison, R. W. Characterization of a monoclonal antibody of the rat liver glucocorticoid receptor. *Endocrinology* **114**, 274–279 (1984).
30. Estes, P. A. *et al.* Immunological analysis of progesterone receptors in human breast cancer. *Biochemistry* **26**, 6250–6262 (1987).

**Acknowledgements.** We thank A. Imbalzano, G. Schnitzler and R. Kingston for hBRG1 antibody; B. Gametchu for the BUGR2 antibody; D. Edwards for the B-30 antibody; J. Torchia and M. Rosenfeld for NcoA1 and NcoA2 antibodies; W. Wang and G. Crabtree for BAFs 60a, 155 and 175 antibodies; S. Goff and B. Strober for hBRG1 plasmids; G. DiMattia and members of our laboratory for critically reading the manuscript; and D. Power for its preparation. This work was supported by grants to T.K.A. from the National Cancer Institute of Canada (NCIC) and the Medical Research Council (MRC) of Canada; T.K.A. is an NCIC Scientist and C.J.F. holds an MRC studentship award.

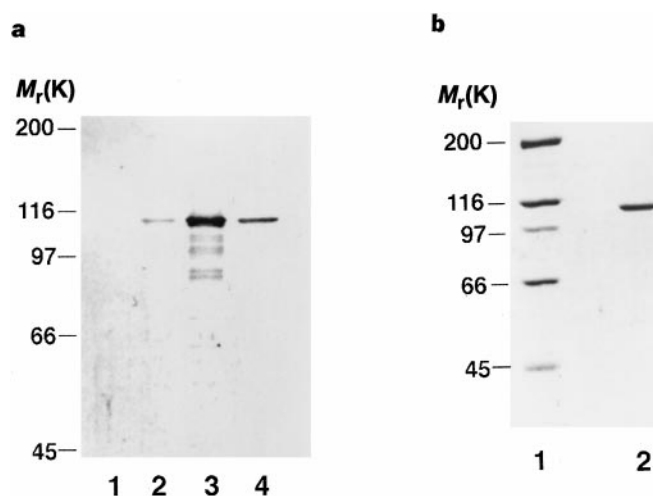
Correspondence and requests for materials should be addressed to T.K.A. (e-mail: tarcher@julian.uwo.ca).

## Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins

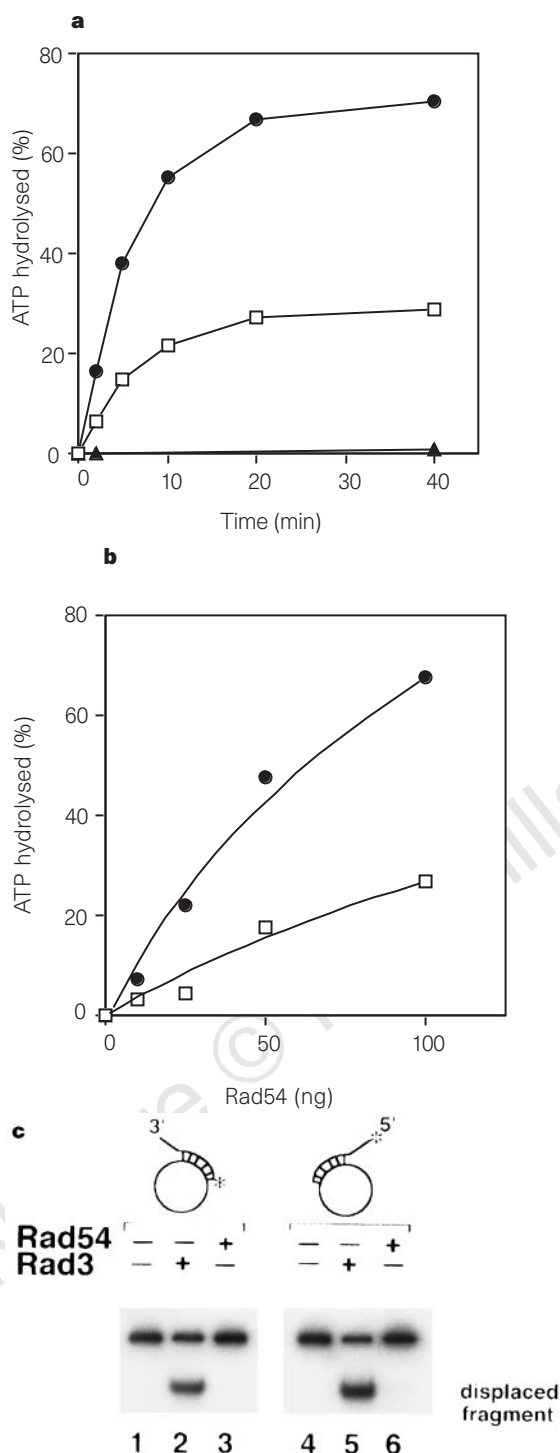
Galina Petukhova, Sabrina Stratton & Patrick Sung

Institute of Biotechnology and Department of Molecular Medicine, University of Texas Health Science Center at San Antonio, 15355 Lambda Drive, San Antonio, Texas 78245, USA

The *Saccharomyces cerevisiae* RAD51 and RAD54 genes are both required for the occurrence of homologous recombination and for the repair of double-stranded DNA breaks<sup>1</sup>. Previous studies have indicated that Rad51 protein, together with the single-stranded DNA-binding factor replication protein A (RPA), can promote the formation of heteroduplex DNA<sup>2–4</sup>, which is a key intermediate in homologous recombination<sup>1</sup>. Here we report the purification of the Rad54 protein to near homogeneity and the biochemical testing of its molecular function. We find that Rad54 protein possesses a double-stranded DNA-dependent ATPase activity, and



**Figure 1** Overexpression and purification of Rad54 protein. **a**, Immunoblot analysis. Extracts from strain BJ5464 (lane 1) or from BJ5464 harbouring plasmid pR54.1 (2  $\mu$  GAL-PGK-6His-RAD54) grown without galactose (lane 2) or with galactose to induce the expression of the Rad54 protein (lane 3), together with 50 ng purified Rad54 protein (lane 4), were immunoblotted with anti-Rad54 antibodies. **b**, An 8% SDS-polyacrylamide gel containing 1  $\mu$ g purified Rad54 protein (lane 2) stained with Coomassie Brilliant Blue.



**Figure 2** Rad54 protein has DNA-dependent ATPase activity. **a**, Purified Rad54 protein (60 ng) was incubated with 1 mM [ $\gamma$ - $^{32}$ P]ATP at 37 °C in 10  $\mu$ l reaction buffer (30 mM K-MOPS, pH 7.2, 3 mM Mg $^{2+}$ , 1 mM DTT, 30 mM KCl) in the absence of DNA (triangles) or in the presence of either 100 ng  $\Phi$ X174 single-stranded DNA (squares) or 200 ng  $\Phi$ X174 double-stranded DNA (circles) for the times indicated. **b**, Different amounts of purified Rad54 protein were incubated with  $^{32}$ P-labelled ATP and single-stranded DNA (squares) or double-stranded DNA (circles) for 8 min, as for **a**. **c**, Rad54 protein has no DNA helicase activity. Either Rad3 protein (90 ng or 100 nM in lanes 2 and 4) or Rad54 protein (150 ng or 147 nM in lanes 3 and 6) was incubated with 5 ng of the two helicase substrates at 37 °C for 30 min, as shown. The hybridized oligonucleotides contained  $^{32}$ P (stars), introduced by treatment with polynucleotide kinase and [ $\gamma$ - $^{32}$ P]ATP. DNA substrates were incubated with Rad54 protein in 10  $\mu$ l buffer R (see Methods) and with Rad3 protein in 10  $\mu$ l buffer as described<sup>9</sup>. Reaction mixtures were analysed as described<sup>9</sup>.

that it interacts with the Rad51 protein. Addition of Rad54 protein to reactions containing Rad51 strongly stimulates the rate of pairing between homologous single-stranded and double-stranded DNA molecules. We conclude that Rad54 acts to overcome kinetic impediments that would limit homologous DNA pairing between recombining chromosomes *in vivo*.

Rad54 protein containing a 6-histidine sequence at its amino terminus was expressed in yeast cells (Fig. 1a), and purified to near homogeneity (Fig. 1b) using a six-step procedure. Rad54 is a member of the Swi2/Snf2 protein family<sup>5</sup> and has sequence motifs suggestive of an ability to bind and hydrolyse ATP<sup>6</sup>. To test for ATPase activity, purified Rad54 protein was incubated with [ $\gamma$ - $^{32}$ P]ATP, and the reaction mixtures were analysed by thin-layer chromatography in polyethylenimine cellulose sheets<sup>7</sup>. However, no significant hydrolysis of ATP ( $\leq 1\%$ ) was observed at pH values from 6.0 to 8.0 (Fig. 2a; data not shown). Because Rad54 may interact with DNA when mediating recombination, we examined whether it would hydrolyse ATP in the presence of DNA. As shown in Fig. 2, the addition of either  $\Phi$ X174 double-stranded or single-stranded DNA resulted in substantial ATP hydrolysis by Rad54, with turnover numbers (molecules of ATP hydrolysed per Rad54 monomer) of  $1.27 \times 10^3 \text{ min}^{-1}$  and  $5.1 \times 10^2 \text{ min}^{-1}$  for double-stranded and single-stranded DNA, respectively. Thus, Rad54 possesses an efficient ATPase activity, but this is dependent on the presence of DNA. Although double-stranded poly(dA)·poly(dT) could activate the ATPase activity of Rad54, neither poly(dA) nor poly(dT) was effective (data not shown), suggesting that the stimulation of Rad54-dependent ATP hydrolysis by  $\Phi$ X174 single-stranded DNA (Fig. 2) was due to secondary structure in the DNA. The use of [ $\alpha$ - $^{32}$ P]ATP as substrate confirmed that the products of hydrolysis are ADP and P<sub>i</sub>. Rad54 protein remained stable for only a short period under the conditions used, causing ATP hydrolysis to reach a plateau after about 20 min incubation (Fig. 2a).

Although helicase-like motifs are present in all of the Swi2/Snf2 family of proteins<sup>5</sup>, none of them has yet been shown to possess DNA helicase activity (see ref. 8 for a review). We therefore investigated whether the Rad54 protein has such activity. Oligonucleotides were hybridized to circular single-stranded pBluescript DNA, to yield two substrates with either a 3' or a 5' tail of 15 bases together with a duplex region of 40 base pairs (Fig. 2c). As a positive control we used the yeast Rad3 protein, a known DNA helicase<sup>9</sup>. As shown in Fig. 2c, while Rad3 protein (100 nM) displaced  $\geq 70\%$  of the hybridized DNA fragment from both substrates, no unwinding of either substrate was observed in the presence of Rad54 protein (147 nM). Rad54 also failed to unwind a partial duplex obtained by hybridizing a 17-base fragment to pBluescript. The addition of Rad51 protein or Rad51 protein filament formed on  $\Phi$ X174 single-stranded DNA<sup>3</sup> to the reaction did not result in unwinding of any of the helicase substrates (data not shown). These results suggest that ATP hydrolysis by Rad54 is not involved in DNA unwinding.

Rad51 protein, a homologue of *Escherichia coli* RecA<sup>10</sup>, mediates pairing of homologous DNA molecules *in vitro*<sup>2-4</sup>. In a model reaction, joint molecules are formed between circular  $\Phi$ X174 bacteriophage plus-strand DNA and linear double-stranded DNA. In the later stages of the reaction, a nicked circular duplex molecule and a displaced linear plus strand are generated as a result of branch migration<sup>2,11</sup>. We examined the effect of Rad54 protein in this model reaction because it interacts with Rad51 protein (Fig. 3a; refs 12 and 13), and genetic evidence suggests that the two proteins function together<sup>1,10</sup>. Because Rad54 protein is much less abundant than Rad51 protein in the cell<sup>12</sup>, a Rad54 protein concentration (94 nM) substoichiometric to that of Rad51 protein (9.3  $\mu$ M) was used in the pairing reaction. As shown in Fig. 3, b and c, addition of Rad54 protein resulted in a strong stimulation of the rate of joint molecule formation, reflecting a more efficient pairing process. A substantial proportion of the joint molecules formed by Rad51 and Rad54 had a slower mobility than those formed in the reaction without Rad54

(Fig. 3b). Considering the stimulatory effect Rad54 exerted on the pairing process, the slower migrating joint molecules were likely to be complex structures consisting of multiply paired single-stranded and double-stranded DNA molecules. At later time points, the amount of nicked-circular duplex and displaced plus-strand DNA formed was lower than that in the control reaction (Fig. 3b). The lower yield of complete DNA strand-exchange products was likely to be due to an impediment to branch migration caused by the presence of multiple single-stranded and double-stranded DNA molecules in the joint molecules complex. It is interesting to note that this type of complex joint molecule is also formed by the UvsX protein, which is the equivalent of RecA in bacteriophage T4 (refs 14, 15).

Rad54 protein itself does not have homologous pairing activity, as no joint molecule was formed in the absence of Rad51. Joint molecule formation was abolished when the reaction did not contain ATP or  $\Phi$ X174 single-stranded DNA, or when the input single-stranded DNA was from the unrelated bacteriophage M13, indicating a requirement for ATP and for DNA homology (data not shown). Previous results showed that efficient homologous pairing is dependent on the single-stranded DNA-binding factor RPA<sup>2-4</sup>. We also found that the omission of RPA from the reaction containing Rad54 protein resulted in a sixfold reduction in product formation (data not shown). Control experiments indicated that the effect of Rad54 protein on homologous pairing could not be attributed to nuclease activity or to a DNA end-joining activity.

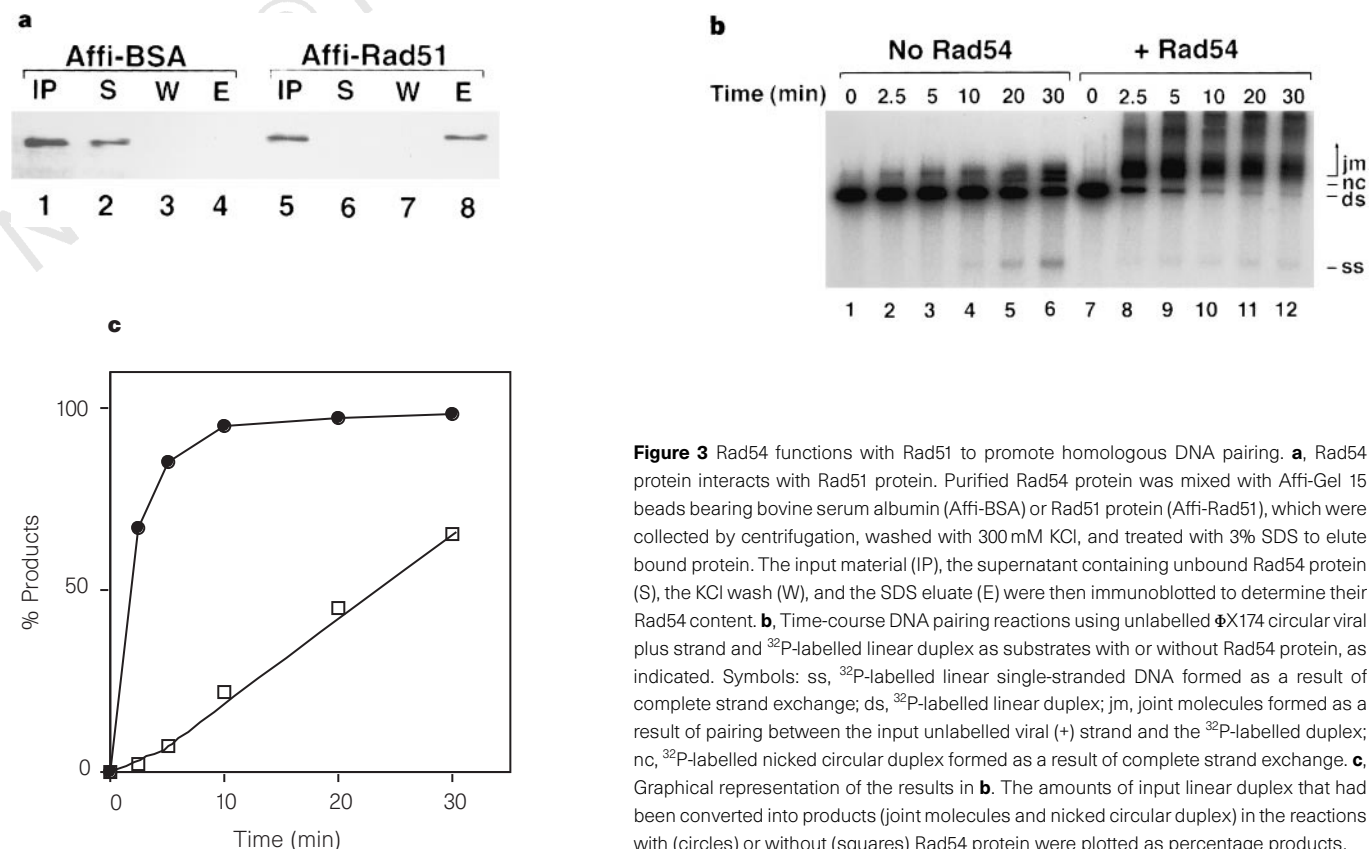
In the double-strand-break model for recombination (see refs 1, 11 for a discussion), the first heteroduplex DNA intermediate formed between recombining DNA molecules is a D-loop. We therefore tested the effect of Rad54 protein on D-loop formation, using  $\Phi$ X174 linear single-stranded DNA and supercoiled DNA as substrates<sup>11</sup>. No joint molecules were seen with Rad51 and RPA alone (Fig. 4a), but the addition of Rad54 protein resulted in efficient joint-molecule formation (Fig. 4a). Densitometric quantification of the gel shown in Fig. 4a revealed that, in the presence of Rad54 protein, 17% of the supercoiled DNA had been converted

into joint molecules after 1 minute of incubation; by the reaction endpoint at 16 minutes, about 75% of the supercoiled DNA had been converted. Heating the reaction mixture at 90 °C for 2 minutes before electrophoresis released both supercoiled DNA and linear single-stranded DNA (Fig. 4a, lane 10), confirming that the joint molecule was indeed a D-loop. Again, D-loop formation requires DNA homology and ATP, and is stimulated by RPA (Fig. 4b).

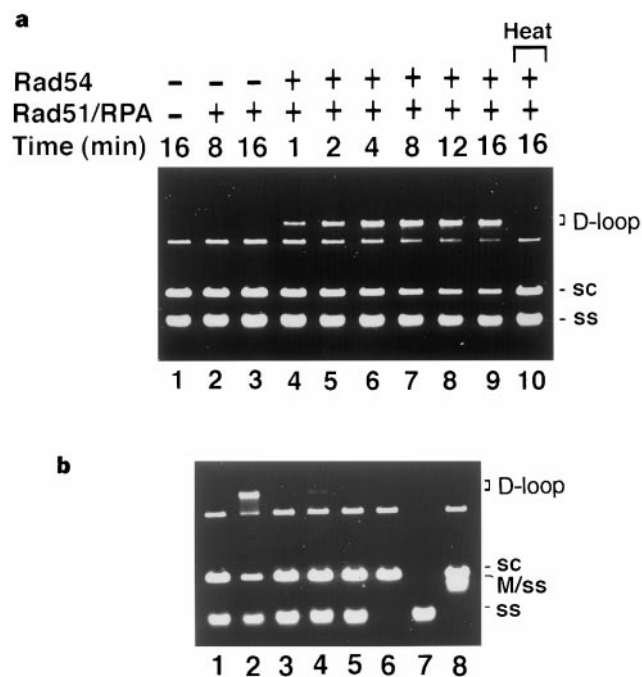
Although Rad51 protein promotes the pairing between a circular single-stranded DNA molecule and its homologous linear double-stranded counterpart<sup>2-4</sup>, the reaction is inefficient in comparison to that catalysed by the prokaryotic homologous-pairing proteins<sup>11,14,15</sup>. Here we have shown that Rad54 protein significantly increases the pairing rate, to a level comparable to that observed with prokaryotic pairing proteins. Rad51 protein alone is incapable of forming a D-loop structure between full-length linear  $\Phi$ X174 single-stranded DNA and supercoiled DNA, although such a DNA structure is likely to represent the first heteroduplex DNA intermediate generated during recombination *in vivo*<sup>11</sup>. We now show that D-loop formation is dependent on Rad54 protein. Rad54 structure and function are conserved among eukaryotes<sup>16,17</sup>, but there does not appear to be a counterpart in prokaryotic organisms. Mutation of the invariant lysine residue (lysine 341) to arginine in the Walker type A nucleotide-binding motif of Rad54 gives a mutant phenotype<sup>13</sup>, suggesting that ATP hydrolysis by Rad54 protein is required for its biological function. Previous studies indicated that homologous DNA pairing occurs within the confines of a Rad51 protein filament formed on single-stranded DNA<sup>3</sup>. Our results suggest a new view of recombination in eukaryotes, in which the Rad51 nucleoprotein filament provides a scaffold upon which catalytic amounts of Rad54 promote homologous DNA pairing at the expense of ATP hydrolysis, by overcoming one or more rate-limiting kinetic barriers. □

# Methods

**Recombination proteins.** A hybrid polypeptide consisting of amino acids 560 to 898 of Rad54 fused to the *E. coli* transcription terminator protein  $\rho$ , purified



**Figure 3** Rad54 functions with Rad51 to promote homologous DNA pairing. **a**, Rad54 protein interacts with Rad51 protein. Purified Rad54 protein was mixed with Affi-Gel 15 beads bearing bovine serum albumin (Affi-BSA) or Rad51 protein (Affi-Rad51), which were collected by centrifugation, washed with 300 mM KCl, and treated with 3% SDS to elute bound protein. The input material (IP), the supernatant containing unbound Rad54 protein (S), the KCl wash (W), and the SDS eluate (E) were then immunoblotted to determine their Rad54 content. **b**, Time-course DNA pairing reactions using unlabelled  $\Phi$ X174 circular viral plus strand and <sup>32</sup>P-labelled linear duplex as substrates with or without Rad54 protein, as indicated. Symbols: ss, <sup>32</sup>P-labelled linear single-stranded DNA formed as a result of complete strand exchange; ds, <sup>32</sup>P-labelled linear duplex; jm, joint molecules formed as a result of pairing between the input unlabelled viral (+) strand and the <sup>32</sup>P-labelled duplex; nc, <sup>32</sup>P-labelled nicked circular duplex formed as a result of complete strand exchange. **c**, Graphical representation of the results in **b**. The amounts of input linear duplex that had been converted into products (joint molecules and nicked circular duplex) in the reactions with (circles) or without (squares) Rad54 protein were plotted as percentage products.



**Figure 4** D-loop formation requires Rad54 protein. **a**, Time-course pairing reaction using  $\Phi$ X174 linear single-strand DNA (ss) and supercoiled DNA (sc) as substrates with or without Rad54 protein, as indicated. In lane 10, the reaction mixture was heated at 90 °C for 2 min before gel electrophoresis to release the linear single-strand DNA and supercoiled DNA from the D-loop. **b**, Homology and ATP requirements. DNA substrates were incubated in buffer (lane 1) and with the combination of Rad51, Rad54 and RPA for 12 min (lane 2). Other lanes were as lane 2, except that Rad51 was omitted in lane 3, RPA was omitted in lane 4, ATP was omitted in lane 5,  $\Phi$ X174 linear single-strand DNA was omitted in lane 6,  $\Phi$ X174 supercoiled DNA was omitted in lane 7, and the  $\Phi$ X174 linear single-strand DNA was replaced by linear single-strand DNA from the unrelated bacteriophage M13 (M/ss) in lane 8.

from *E. coli*, was used for raising a polyclonal antiserum in rabbits. Anti-Rad54 antibodies were purified by affinity chromatography<sup>9</sup>. The *RAD54* gene containing a sequence encoding six His residues at the amino terminus was fused to the *GAL-PGK* promoter to yield plasmid pR54.1 (2 $\mu$ , *LEU2-d*, *GAL-PGK-6His-RAD54*), which was introduced into the yeast strain BJ5464 (*Mata*, *ura3-52*, *trp-1*, *leu2Δ1*, *his3Δ200*, *pep4::HIS3*, *prbΔ1.6R*). Overnight yeast cultures were diluted tenfold into fresh synthetic medium lacking leucine and containing 2% galactose, and grown at 30 °C for 12 h. For the purification of Rad54 protein, the extract was prepared<sup>9</sup> and precipitated with ammonium sulphate, then chromatographed on Q-Sepharose, SP-Sepharose, hydroxyapatite, Mono-S, and nickel NTA-agarose (Qiagen). We obtained 60  $\mu$ g of nearly homogeneous 6His-Rad54 protein from 200 g of yeast paste. The purified Rad54 protein was stored in 20 mM  $\text{KH}_2\text{PO}_4$ , pH 7.4, 10% glycerol, 0.5 mM EDTA, 0.5 mM DTT, 300 mM KCl. The details will be described elsewhere. Rad51 protein and RPA were purified from yeast as described<sup>2</sup>.

**Binding of Rad54 protein to Rad51 Affi-gel.** Purified Rad54 protein, 1  $\mu$ g in 150  $\mu$ l buffer T (20 mM Tris-HCl, pH 7.5, 10% glycerol, 0.5 mM EDTA, 0.5 mM DTT, and 0.01% NP-40) containing 150 mM KCl and 200  $\mu$ g ml<sup>-1</sup> BSA, was mixed with 7.5  $\mu$ l of Affi-gel 15 beads containing covalently conjugated Rad51 protein (3.2 mg ml<sup>-1</sup>) or BSA (5 mg ml<sup>-1</sup>) for 45 min at 25 °C. The beads were collected by a 5-s centrifugation in a microfuge, washed with 300  $\mu$ l buffer T containing 300 mM KCl, before being treated with 40  $\mu$ l of 3% SDS at 37 °C for 10 min to elute bound Rad54 protein. The input material (10  $\mu$ l), the supernatant that contained unbound Rad54 (10  $\mu$ l), the KCl wash (10  $\mu$ l), and the SDS eluate (3  $\mu$ l) were immunoblotted to determine their Rad54 content.

**Homologous DNA pairing reactions.** (1) System that uses circular viral plus strand and linear duplex. The reaction used in Fig. 3b (50  $\mu$ l final volume) was

assembled by mixing 20  $\mu$ g Rad51 protein (9.3  $\mu$ M) in 4  $\mu$ l storage buffer with 325 ng  $\Phi$ X174 viral plus-strand DNA (25  $\mu$ M nucleotides) in 4  $\mu$ l TE (10 mM Tris-HCl, pH 7.5, 0.2 mM EDTA), in 40  $\mu$ l of buffer R (35 mM potassium MOPS, pH 7.2, 30 mM KCl, 2 mM ATP, 3 mM  $\text{MgCl}_2$ , 1 mM DTT, and an ATP-regenerating system consisting of 20 mM creatine phosphate and 1.2  $\mu$ g creatine kinase). After a 5-min incubation at 37 °C, 6.5  $\mu$ g RPA (1.1  $\mu$ M) in 2  $\mu$ l storage buffer was added, followed by a 5-min incubation at 37 °C, and then 480 ng Rad54 protein (94 nM) in 1  $\mu$ l storage buffer, 200 ng of 5' <sup>32</sup>P-labelled linear  $\Phi$ X174 double-stranded DNA (15.4  $\mu$ M nucleotides) in 3  $\mu$ l TE, and 4  $\mu$ l 50 mM spermidine (4 mM) were added. The complete reaction mixture was incubated at 37 °C, and 6- $\mu$ l portions were withdrawn at the times indicated and processed for agarose-gel electrophoresis as described<sup>2,3</sup>.

(2) D-loop assay. Linear  $\Phi$ X174 single-strand DNA was obtained by hybridizing a 20-mer oligonucleotide to the viral plus strand to create a *Pst*I site, followed by complete digestion with *Pst*I. The linearized single-strand DNA was purified with a GeneClean kit (Bio 101). The reaction used for Fig. 4a (50  $\mu$ l final volume) was assembled by mixing 20  $\mu$ g Rad51 protein (9.3  $\mu$ M) in 4  $\mu$ l storage buffer with 325 ng of  $\Phi$ X174 linear single-strand DNA strand (25  $\mu$ M nucleotides) in 4  $\mu$ l, in 40  $\mu$ l buffer R. After a 5-min incubation at 37 °C, 6.5  $\mu$ g RPA (1.1  $\mu$ M) in 2  $\mu$ l storage buffer was added, followed by a 5-min incubation at 37 °C. Next, 480 ng Rad54 protein (94 nM) in 1  $\mu$ l storage buffer, and 240 ng  $\Phi$ X174 double-stranded DNA (18.5  $\mu$ M nucleotides; about 85% supercoiled form) in 3  $\mu$ l TE and 4  $\mu$ l of 50 mM spermidine (4 mM) were incorporated. The reaction mixture was incubated at 37 °C, and 7.5  $\mu$ l portions were withdrawn at the times indicated and processed for agarose gel electrophoresis<sup>2,3</sup>. DNA species were visualized by staining with ethidium bromide (3  $\mu$ g ml<sup>-1</sup> in H<sub>2</sub>O) for 2 h, and gels were destained for 3 h in water before being photographed using Polaroid type 55 films. Image analysis of the photographic negative was done using a Molecular Dynamics SI densitometer. For Fig. 4b, the reaction mixtures were assembled in the same manner, except that they were scaled down four times. The linear single-strand M13 DNA used in Fig. 4b was obtained by hybridizing an 18-mer oligonucleotide to the circular viral plus strand of M13 mp18 to create a *Bgl*II site, followed by digestion of the DNA with *Bgl*II and purifying the product using a GeneClean kit.

Received 10 November 1997; accepted 9 February 1998.

- Petes, T. D., Malone, R. E. & Symington, L. S. in *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis, and Energetics* 407–521 (Cold Spring Harbor Laboratory Press, 1991).
- Sung, P. Catalysis of ATP dependent homologous DNA pairing and strand exchange by the yeast Rad51 protein. *Science* **265**, 1241–1243 (1994).
- Sung, P. & Roberson, D. L. DNA strand exchange mediated by a Rad51-ssDNA nucleoprotein filament with polarity opposite to that of RecA. *Cell* **82**, 453–461 (1995).
- Sugiyama, T., Zaitseva, E. M. & Kowalczykowski, S. C. A single-strand DNA binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *J. Biol. Chem.* **272**, 7940–7945 (1997).
- Eisen, J. A., Sweder, K. S. & Hanawalt, P. C. Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res.* **23**, 2715–2723 (1995).
- Emery, H. S., Schild, D., Kellogg, D. E. & Mortimer, R. K. Sequence of *RAD54*, a *Saccharomyces cerevisiae* gene involved in recombination and repair. *Gene* **104**, 103–109 (1991).
- Randerath, K. & Randerath, E. Ion-exchange chromatography of nucleotides on poly-(ethyleneimine)-cellulose thin layers. *J. Chromatogr.* **16**, 111–125 (1964).
- Pazin, M. J. & Kadonaga, J. T. Swi2/Snf2 and related proteins: ATP driven motors that disrupt protein-DNA interactions? *Cell* **88**, 737–740 (1997).
- Sung, P., Matson, S. W., Prakash, L. & Prakash, S. Rad3 protein of *Saccharomyces cerevisiae* is a DNA helicase. *Proc. Natl Acad. Sci. USA* **84**, 6045–6049 (1987).
- Shinohara, A. & Ogawa, T. Homologous recombination and the roles of double-strand breaks. *Trends Biochem. Sci.* **20**, 387–391 (1995).
- Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. Biochemistry of homologous recombination in *E. coli*. *Microbiol. Rev.* **58**, 401–465 (1994).
- Jiang, H. et al. Direct association between the yeast Rad51 and Rad54 recombination proteins. *J. Biol. Chem.* **271**, 33181–33186 (1996).
- Cleaver, B. et al. Recombinational repair in yeast: functional interactions between Rad51 and Rad54 proteins. *EMBO J.* **16**, 2535–2544 (1997).
- Formosa, T. & Alberts, B. M. Purification and characterization of the bacteriophage T4 *UvsX* protein. *J. Biol. Chem.* **261**, 6107–6118 (1984).
- Salinas, F. & Kodadek, T. Phage T4 homologous strand exchange: a DNA helicase, not the strand transferase, drives polar branch migration. *Cell* **82**, 111–119 (1995).
- Bezzubova, O., Silbergliet, A., Yamaguchi-Iwai, Y., Takeda, S. & Buerstedde, J. M. Reduced X-ray resistance and homologous recombination frequencies in a *RAD54*–/– mutant of the chicken DT40 cell line. *Cell* **89**, 185–193 (1997).
- Esser, J. R. et al. Disruption of mouse *RAD54* reduces ionizing radiation resistance and homologous recombination. *Cell* **89**, 195–204 (1997).

**Acknowledgements.** This work was supported by a PHS grant from the National Institute of Environmental Health Sciences.

Correspondence and requests for materials should be addressed to P.S. (e-mail: sung@uthscsa.edu).



# Mastering the molecular domain

**Molecular biologists converge on Washington, DC for the annual American Society for Biology and Molecular Biology meeting, May 17–21. On exhibition will be tools for PCR, purification, sequencing, and bibliographic support.**

## GS-2000 sequencer

From Web Scientific

*An automated DNA sequencer and fragment analysis system, developed by Corbett*

Designed for high performance and low cost, this instrument uses ultra-thin, 80- $\mu$ m gels, which are said to rival the performance of capillary systems, but allow greater numbers of samples to be run simultaneously. According to the manufacturer, the gel pouring is bubble free and simple, and the same gel can be used up to five times. Gels can be quickly loaded in either 48- or 64-well formats, using a standard multi-channel pipette (0.4-mm wells obviate the need for special pipette tips). This was designed to be an ultra-fast, automated fluorescence system, which is said to read up to 400 bases in 3 h 10 min, with high sensitivity and resolution. The instrument can detect and size 100 attomoles of fluorescently labelled DNA in a single peak, and works with most commercially available fluorophores. DNA fragments can be sized to  $\pm 0.2$  nt across the entire gel. For RFLPs and RAPDs under non-denaturing conditions, ethidium bromide can be used to detect as little as 500 attomoles of DNA. Genotyping runs can be carried out in less than a stated 30 min. With the addition of a cooling accessory, the instrument can also carry out SSCP, says Web Scientific.

**Reader Enquiry No. 100**

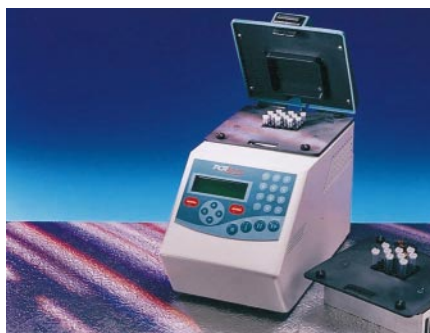
## SQL\*GT

From Perkin-Elmer

*A genetic tracking system that supports bioinformatics sample tracking*

Based on PE Nelson's laboratory information management system (LIMS), this program has been combined with two complementary molecular informatics products: the BioLIMS system for DNA sequence and results management, and the BioMerge system for genomics data analysis and integration. It was created to perform project and sample management associated with genetic analysis. SQL\*GT helps automate sample tracking, manage workflow and integrate information with existing database applications. The plate- and sample-tracking architecture allows immediate determination of library or sample location. The control software uses a graphical plate interface, and interacts directly with the ABI Prism XL automated DNA analysis system by creating sample sheets for gel loading.

**Reader Enquiry No. 101**



Make a run on DNA with Hybaid's PCR Sprint.

## Amplification

### PCR Sprint

From Hybaid

*A new oil-free personal thermal cycler*

This unit uses a screen and keypad layout to allow rapid, easy programming, with the same easy-to-use programming and temperature control characteristics of the company's larger capacity instruments. Operators are guided through a series of large, user friendly screens. Hybaid's 'active-tube' control and 'simulated-tube' control systems are standard features on this system. These are said to allow more accurate and reproducible temperature control than is achievable on thermal cyclers that rely on measurement of sample block temperature to control the cycling. This system also features robust Peltier block technology and is said to be compact and affordable.

**Reader Enquiry No. 102**

## BioFocus LIF<sup>2</sup>

From Bio-Rad

*A new fluorescence detector for PCR analysis, which is stated to be  $\times 1000$  more sensitive than ultraviolet detection*

The increased sensitivity of this new laser-induced fluorescence (LIF) detector is said to increase speed and accuracy of PCR product measurements. This method of nucleic acid fragment analysis uses capillary electrophoresis, which is said to reduce the number of DNA amplification cycles that are needed. PCR products, restriction fragments, microsatellites and mutations can be detected at stated concentrations of 0.5  $\mu$ M or less. BioFocus LIF<sup>2</sup> incorporates both a 488-nm Ar laser and a 594-nm He/Ne laser, which can be set up with one or two data-collection channels. In the two-channel configuration, the sample can be run simultaneously with an internal standard, which helps to eliminate run-to-

run variations. The instrument should prove especially useful in conjunction with the company's BioFocus dsDNA low-background fluorescent stain kit, detecting DNA fragment concentrations down to  $8 \times 10^{-12}$  M. In addition, the detector provides automatic real-time data integration and report printing.

**Reader Enquiry No. 103**

## RNase-free water

From Severn Biotech

*RNase-free, DEPC-treated water specifically for PCR reactions*

Supplied in a 5-ml amber vial, this product should prove useful to those doing PCR, but would like the increased security of eliminating water as a potential source of contamination in enzymatic reactions. The volume is said to be adequate for most applications and can be disposed of after use, eliminating cross-contamination by other workers. Each vial contains water that has been bacterial filtered, purified to 18 M $\Omega$ , treated with DEPC and autoclaved. (This product is sold in packs of 100.)

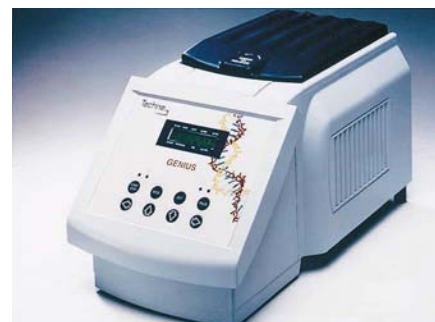
**Reader Enquiry No. 104**

## Genius thermal cycler

From Techne

*A 384-well thermal cycler for large-capacity, high-throughput sequencing and PCR*

The 384 block is interchangeable with all the other blocks available for Genius, allowing the easy scale-up and transfer of protocols. Using a 384-block format allows the researcher to scale up from existing 96-well formats and to reduce costs by reducing sample volume and reagent usage. The cycler is said to control the temperature accurately, up to 99 °C heating and down to 4 °C cooling, with 0.5 °C block uniformity. Each unit has a self-adjusting heated lid to allow for oil-free reactions. This unit is designed for easy programming and will accept all of the



'Brains and brawn' with the Genius thermal cycler.

current complex protocols (step down, nesting, long, rapids, and the like).

**Reader Enquiry No. 105**

## Separations

### IPGphor

From Amersham Pharmacia Biotech

*An isoelectric focusing (IEF) system for two-dimensional electrophoresis*

This system is designed for high-resolution, first-dimension protein separations on immobilized, pH-gradient gels. It offers simplified sample handling and is said to reduce the number of IPG-strip manipulations, as well as reducing separation times without a loss in resolution. The system includes a programmable 8,000-V, 1.5-mA power supply and solid-state Peltier temperature control (from 18–25 (±1) °C) for up to twelve IPG-strip separations at one time. A ceramic strip holder functions in both the rehydration and separation steps. In the strip holder, an IPG strip is rehydrated gel-side down in a small volume of rehydration solution. Near the ends of the rehydration channel, the gel rests on platinum electrodes, which extend through the base of the strip holder to external contact points. By placing the holder on the contact areas of the platform, the holder simultaneously makes contact with the power supply and the temperature control plate. Thus, the IEF step is carried out in the same strip holder as rehydration, without moving the strip or the holder. The system controller can store nine programmes, each with up to nine stepped or ramped voltage changes. Users can also use a delayed start, loading the system in the afternoon and having IEF start automatically during the night, for example.

**Reader Enquiry No. 106**

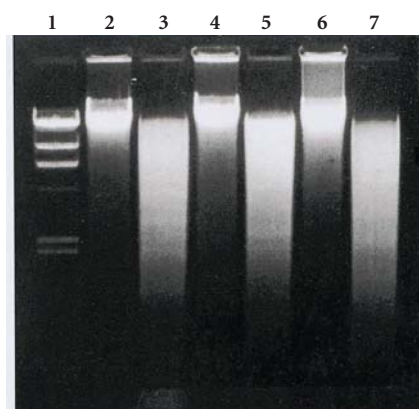
### Centrifugal filter tubes

From Eppendorf

*A range of filter inserts with their respective centrifuge tubes*

Created for concentration, purification or desalting of various samples, these filter tubes are available in three volumes (0.5 ml, 4.0 ml and 15 ml), and in six different cut-off sizes, each ranging from 4K to 0.45 µm. This covers all relevant applications, ranging from the concentration of peptides all the way up to bacteria and cell harvesting. The membranes are characterized by extremely low affinity for protein or nucleic acids and thus correspondingly high recovery rates, says Eppendorf. The special construction of the inserts ensures reliable sealing of the tubes during centrifugation while, at the same time, maximizing concentration and giving defined retention volumes. Eppendorf thus offers an efficient solution for rapid concentration of liquid samples.

**Reader Enquiry No. 107**



Genomic DNA isolated with DNAzol ES.

### DNAzol ES

From Molecular Research Center

*A ready-to-use reagent for the isolation of genomic DNA from plant material*

The reagent contains a guanidine detergent, lysing mixture that is designed to facilitate plant DNA extraction while optimizing the removal of contaminating factors. To accommodate the wide variety of plants studied, the protocol includes optional steps appropriate for more difficult plant tissues. One gram of plant tissue requires only 3 ml of DNAzol ES for DNA extraction. In the figure (above), DNA has been isolated from tomato, philodendron and green pepper leaves (lanes 2–3, 4–5 and 6–7, respectively). DNA from lane 3, 5 and 7 was restricted with 5 U *EcoRI* per mg of DNA, 4 h at 37 °C and electrophoresed in one per cent agarose.

**Reader Enquiry No. 108**

## General software

### Grant forms

From ScienceMedia

*Software for writing NIH grant proposals*

The new software can be downloaded and reviewed at <<http://www.grantforms.com/>>

**Reader Enquiry No. 109**

### EndNote 3.0

From Niles Software

*An Internet search client that searches remote databases and creates bibliographies*

This new version allows users to search remote bibliographic databases on the Internet and includes better integration with electronic publications. Currently, researchers must go through a multiple-step process to download references from online databases and import them into their personal bibliographic database. EndNote 3.0 has been written to simplify this multiple-step process by integrating the two functions into one program. The new program has more than 100 connection files that will link to resources such as university card catalogues, the Library of Congress and to more specialized data-

bases, such as MEDLINE, PsycInfo and ERIC. The user simply opens the connection file and can begin searching, just as they would in their own EndNote database. Then selected references can be simply dragged and dropped into their local EndNote database. The new version also allows the user to launch a web browser from within the application, directly to the URL.

**Reader Enquiry No. 110**

*These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.*

## ADVERTISEMENTS

### ANTIBODIES

#### For Research & Routine Histology

Inflammation – Angiogenesis – Tumors  
Leukocytes – Endothelial Cells  
Pathological Markers

#### Custom Monoclonals and Polyclonals

BMA Biomedicals Ltd.  
(800) 634-3183 (US)  
Intl: +41 61 811 62 22  
+41 61 811 60 06 (fax)  
[info@bma.ch](mailto:info@bma.ch)

READER ENQUIRY NO. 28

NEWCASTLE UNIVERSITY  
MOLECULAR BIOLOGY UNIT



### PROTEIN SEQUENCING

*Membrane-bound, solid or solution samples*

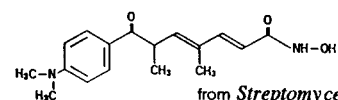
**£12 per cycle**

5 cycles minimum - no set up charge

For further information contact Dr. Joe Gray  
Phone: +44 191 222 8612  
Fax: +44 191 222 8129  
Email: [JOE.GRAY@newcastle.ac.uk](mailto:JOE.GRAY@newcastle.ac.uk)

READER ENQUIRY NO. 14

### Trichostatin A



from *Streptomyces*

Catalog No. 204-11991

**This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call**  
**Wako BioProducts: (800) 992-9256; (804) 271-7677.**

READER ENQUIRY NO. 98